NAVAL POSTGRADUATE SCHOOL
Monterey, California

DISSERTATION

THEORY OF MULTIRATE STATISTICAL SIGNAL PROCESSING AND APPLICATIONS

by

Ryan J. Kuchler

September 2005

Dissertation Supervisor: Charles W. Therrien

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave blank*)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 2005	Doctoral Dissertation

4. TITLE AND SUBTITLE Theory of Multirate Statistical Signal Processing and Applications	5. FUNDING NUMBERS
6. AUTHORS Kuchler, Ryan J.	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey CA 93943-5000	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION/AVAILABILITY STATEMENT	12b. DISTRIBUTION CODE
Approved for public release; distribution is unlimited.	

13. ABSTRACT(*maximum 200 words*)

This dissertation develops basic theory and applications of statistical multirate signal processing. Specific tools and terminology for describing multirate systems in the time and frequency domains are presented. An optimal multirate estimator is derived in both a direct form and recursive form. The recursive form of the optimal estimator allows calculation of the relative change in performance when input signals are added or removed from the multirate system. The optimal multirate filtering problem also is specialized to the case of optimal multirate linear prediction. An efficient method for calculating the multirate linear prediction coefficients and error variances is developed through the use of the multichannel Levinson recursion and generalized triangular UL factorization. Finally, a multirate sequential classifier is derived and applied to the problem of target classification. It is shown that classifier parameters needed for implementing the multirate sequential classifier are the same as those for multirate linear prediction. The methods presented in this dissertation are useful for multisensor fusion, particularly when the sensors are operating at different rates.

14. SUBJECT TERMS Multirate Statistical Signal Processing, Linear Estimation, Linear Prediction, Classification	15. NUMBER OF PAGES 195
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500 Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18 298-102

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

THEORY OF MULTIRATE STATISTICAL SIGNAL PROCESSING AND APPLICATIONS

Ryan J. Kuchler
Lieutenant Commander, United States Navy
B.S., United States Naval Academy, 1993
M.S., Naval Postgraduate School, 2002

Submitted in partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY IN ELECTRICAL ENGINEERING

from the

NAVAL POSTGRADUATE SCHOOL
September 2005

Author: Ryan J. Kuchler

Approved by: Charles W. Therrien, Professor of Electrical Engineering
Dissertation Supervisor and Committee Chair

Roberto Cristi Professor of Electrical Engineering	Murali Tummala Professor of Electrical Engineering
Carlos F. Borges Associate Professor of Mathematics	Anthony J. Healey Professor of Mechanical Engineering

Approved by: Jeffrey B. Knorr, Chair, Department of Electrical and Computer Engineering

Approved by: Knox T. Millsaps, Associate Provost for Academic Affairs

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

This dissertation develops basic theory and applications of statistical multirate signal processing. Specific tools and terminology for describing multirate systems in the time and frequency domains are presented. An optimal multirate estimator is derived in both a direct form and recursive form. The recursive form of the optimal estimator allows calculation of the relative change in performance when input signals are added or removed from the multirate system. The optimal multirate filtering problem also is specialized to the case of optimal multirate linear prediction. An efficient method for calculating the multirate linear prediction coefficients and error variances is developed through the use of the multichannel Levinson recursion and generalized triangular UL factorization. Finally, a multirate sequential classifier is derived and applied to the problem of target classification. It is shown that classifier parameters needed for implementing the multirate sequential classifier are the same as those for multirate linear prediction. The methods presented in this dissertation are useful for multisensor fusion, particularly when the sensors are operating at different rates.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

EXECUTIVE SUMMARY

Sensor fusion is an important component of target detection and classification. The ability to exploit correlated information among multiple sensors with different operating characteristics can lead to improved detection and classification performance. This dissertation addresses multirate sequential classification through the development of a common multirate framework and the multirate linear prediction equations.

In order to lay the foundation for representing multirate signals, key terms applicable to multirate systems are defined. These are the fundamental rate, system period, system phase and decimation factors. These terms are used to explicitly describe the periodic nature of the multirate system and allow for the appropriate selection of values for the periodic components in the multirate system. In addition, fundamental building blocks of the multirate system are presented, along with the statistical characterizations of multirate signals as they pass through the multirate building blocks. In particular, a linear periodically time-varying filter is presented that permits the input and output signals to be sampled at different rates.

The first multirate optimal filter considered in this dissertation is the optimal linear estimator. The explicit direct form that estimates multiple input signals at possibly different sampling rates is first presented followed by a recursive *innovations* form. This innovations form of the optimal estimator separates the direct form optimal filters into modified optimal filters and cross filters. These cross filters remove any information from one signal that is contained in the other signals, in an ordered fashion. In essence, a new set of input signals that are mutually orthogonal are derived and used as the inputs to the modified optimal filters. Using the innovations form of the optimal filter allows one to calculate the relative change in performance in the optimal estimator when input signals are added or removed from the system.

The optimal filtering problem is specialized to the case of optimal multirate linear prediction. The multirate Normal equations are derived for a multirate system with multiple input and output signals which are observed at different sampling rates. For a multirate system with a system period K, there are up to K distinct sets of prediction coefficients and error covariance matrices that apply in a periodic fashion. An efficient method for calculating the multirate linear prediction coefficients and error variances is developed through the use of the multichannel Levinson recursion and generalized triangular UL factorization.

Finally, a multirate sequential classifier is derived starting from the basic theory of sequential hypothesis testing. It is shown that classifier parameters needed for implementing the multirate sequential classifier are the same as those for multirate linear prediction. A multirate sequential classifier is then implemented and tested using audio files of a propellor plane and three A-10 jet aircraft. The experiments tested the classifier performance in selecting between the propellor plane and jet aircraft as certain system parameters are changed. These parameters are the signal-to-noise ratios of the observed signal and the length of the training data. In addition, the performance of the multirate classifier is compared to that of the single-channel and multichannel classifiers using similar data.

ACKNOWLEDGMENTS

The successful completion of my doctoral program and this dissertation is the result of contributions of many individuals. I have been privileged to have the continuous support of relatives, friends, and colleagues to whom I express my most heartfelt gratitude. Special thanks are owed the following individuals.

First and foremost my parents, Jim and Laura, who have always provided encouragement and support in all my endeavors.

My committee members - Dr. Charles Therrien, Dr. Carlos Borges, Dr. Roberto Cristi, Dr. Anthony Healey and Dr. Murali Tummala - who have helped to make this endeavor a most rewarding experience. I would like to give special recognition and my sincerest appreciation to Dr. Charles Therrien for his mentoring, guidance and patience as my committee chairman and dissertation supervisor.

My high school calculus and electronics teachers, Mr. Andrew Pasquale and Mr. Bernard DiNatale, for instilling in me a love of mathematics and engineering. Without their teaching and encouragement I would never have had this opportunity.

All my friends at the Carmel Mission, who have been like family to me and have helped to make my time at school and outside of school most enjoyable and memorable. I especially would like to thank the Carmel Mission Choir, the choir director Ms. Kathy Anderson, and the organist Mr. Ed Soberanes, who warmly accepted me as one of their own, and the priests, Fr. John Griffin and Fr. Emil Robu, who provided me wise counsel during this arduous endeavor.

And finally, all my friends at A Taste of Monterey, who graciously provided a unique, reflective environment for me to study, as well as many opportunities for me to learn about some finer things in Monterey, like the wine, in addition to learning engineering.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

Within the field of signal processing, a growing interest in the area of multirate signal processing has developed over the past three decades. This area focuses on the processing of systems that contain multiple signals that can occur at different sampling rates. Since this area deals with systems that have components operating at different sampling rates, processing techniques and descriptions are needed to account for any disparity between these sampling rates. Many advantages of multirate signal processing have been found, and techniques have been applied to many areas, such as telecommunications, digital audio encoding/decoding, speech and image processing, and geophysical signal processing [Ref. 1, 2, 3].

Most of the research on multirate signal processing, beyond development and characterization of the basic building blocks, has centered on filter bank theory and multirate Kalman filtering [Ref. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15]. A brief discussion on each of these areas is provided below. In the majority of the filter bank research, there is one input signal (or vector of input signals sampled at the same rate). Within the filter bank structure, the signal is separated into different subbands sampled at different rates. Signal processing techniques are applied, and the separate signals are then synthesized into one output signal (or vector of output signals sampled at the same rate). This area of research has greatly improved the knowledge and understanding of single-input single-output (SISO) and multiple-input multiple-output (MIMO) filter bank structures; however, it has not developed methods for processing multiple input signals at different sampling rates.

In the area of statistical multirate signal processing, there are a number of papers published. One area that these papers have focused on is the characterization of periodic random processes. These papers present methods to describe periodic random processes in both the time and frequency domains and discuss such key concepts as cyclostationarity of periodic random processes. Another significant area

of research in statistical multirate signal processing is the development of state-space multirate Kalman filters. These Kalman filters have been applied to a variety of systems, such as small aircraft orbit control, helicopter passive ranging and machinery control. In addition, multirate Kalman filter research has been applied to tracking and estimation of mobile robots as well as recovery of lost speech packets in speech signal processing.

This dissertation approaches statistical multirate signal processing from the view point of optimal filtering (i. e., Wiener filtering) of multiple input signals at different sampling rates. Following the investigation of statistical characteristics for multirate signals, an optimal estimator for a desired output sequence is developed. This development is then generalized to the linear prediction problem, which leads to the general form of the optimal multirate Wiener filter. Finally, this optimal filter is applied to the sequential target classification problem. Much of the foundation for the multirate signal processing in this dissertation is derived from [Ref. 16, 17]. In addition the optimal estimation and prediction problems in this dissertation are multirate extensions of analogous single-channel and multichannel concepts that can be found in [Ref. 18, 19]. The multirate sequential classification algorithm is derived from previous work on the single-channel and multichannel classification theory and algorithms of [Ref. 20, 21, 22].

The following paragraphs review some of the research that has been conducted in multirate signal processing. This includes multirate theory, multirate filter bank theory, statistical signal characterization and multirate Kalman filtering. It is by no means exhaustive, and indeed much work is still being conducted today.

A. LITERATURE REVIEW

The interest in multirate signal processing gained much popularity following the 1975 IEEE Arden House Workshop for Acoustics, Speech, and Signal Processing

[Ref. 16]. Much of the initial research in multirate signal processing has focused on how to describe the multirate system, including the descriptions of some fundamental building blocks of the multirate system and on how to apply multirate signal processing to filter bank theory. In 1983, Crochiere and Rabiner published the first comprehensive book on multirate signal processing [Ref. 16]. Their work presents the basics of sampling rate conversion through decimation and expansion, including time and frequency characterization of multirate signals. The application that they primarily focus upon is improving efficiency in a multirate system with large sampling rate changes through multistage structures. These multistage structures break the sampling rate changes into multiple stages, which can allow relaxation of filter design criteria. In addition to multistage applications, Crochiere and Rabiner also describe several forms of uniform filter banks (e.g., the discrete Fourier transform (DFT) filter bank and the uniform single-sideband (SSB) filter bank) where the signal is separated into multiple parts and all parts are decimated by the same value. They extend the DFT filter bank to a generalized non-uniform DFT filter bank, which allows the separated signals to be decimated by different values.

Ten years after the book by Crochiere and Rabiner, the next major publication in multirate signal processing was by Vaidyanathan in 1993 [Ref. 17]. Vaidyanathan's work addresses the basics of multirate signal processing but provides greatly expanded applications in filter bank theory. Vaidyanathan presents and thoroughly describes many types of multirate filter banks, including maximally decimated filter banks, paraunitary perfect reconstruction filter banks, linear phase perfectly reconstruction quadrature mirror filter (QMF) banks and cosine modulated filter banks. His work also presents the concept of periodically time-varying filters, which plays an important part in the processing of multirate signals and the problem of quantization error. Lastly, Vaidyanathan introduces the relationship between wavelet transform theory and multirate signal processing since the wavelet transform inherently uses nonuniform decimation in developing its subbands. Both of these books [Ref. 16, 17] have

provided the foundation on which much of the research on multirate signal processing since then has been built.

In addition to these two seminal volumes on multirate signal processing, there have been many contributions in journal papers and workshop and conference publications. Shenoy, Burnside and Parks extend previous work on multirate filter bank theory to optimum minimax filters [Ref. 1]. By using a generalized Fourier analysis, they derived a new error criteria for multirate filter design, which can be used to design the optimum minimax multirate filters for a specific input signal class. In addition, Chen and Vaidyanathan [Ref. 4] extend the concept of polyphase filters to describe rational sampling rate alterations. These polyphase filters are useful in representing perfect reconstruction properties of multidimensional delay-chain systems and periodicity properties of decimated periodic signals. Rules for multirate structures were also extended by Evans, Bamberger and McClellan [Ref. 5] to multidimensional structures by finding greatest common subblattices and computing coset vectors.

Since the multirate system contains signals at different sampling rates, there is a natural periodicity that can be seen in the multirate system. Some basic forms of linear periodically time-varying (LPTV) filters were described by Vaidyanathan; however, Saadat Mehr and Chen [Ref. 23] present two methods of describing LPTV structures, each of which consists of a periodic switch connected to several linear time-invariant (LTI) systems. Saadat Mehr and Chen show that these structures can be used to solve a general approximation problem where an LPTV system with period p is approximated by an LPTV system with period $\hat{p}$. They extend their work on LPTV structures to address polyphase and alias-component representations of LPTV systems [Ref. 24]. They show that in general a filter bank can be represented by two LPTV systems connected in a cascading manner. They also show that a p-channel filter bank with period m can be represented by an mp-channel LTI filter bank if p and m are relatively coprime integers.

Other research on time-varying systems has included work by Phoong and Vaidyanathan [Ref. 25] on using a polyphase approach to filter banks to address unusual properties not exhibited by LTI filter banks. This work is extended in [Ref. 26] to discuss the factorability of linear time-varying lossless filter banks. In particular, they show that all degree-one lossless linear time-varying (LTV) systems can be decomposed into a time-dependent unitary matrix and a lossless dyadic-based LTV system. Vaidyanathan expanded his work on periodic systems to address periodic systems with allpass and paraunitary properties [Ref. 27]. By providing a state-space representation of periodic LTI systems and introducing the concept of reachability and observability in terms of multirate periodic systems, Vaidyanathan shows that reachability and observability are not related to the system minimality in a simple way, unlike traditional state-space linear systems. In [Ref. 28], Gadre and Patney address issues associated with aliasing cancelation and perfect reconstruction within a vector context, and they define a vector multirate system. These definitions lead to conditions whereby a vector LPTV system can become a time-invariant system. Further, Spurbeck and Scharf [Ref. 29] applied spectral factorization techniques to the filter design for periodically correlated time series.

Another approach to representing periodic systems was introduced by Misra [Ref. 30]. Misra shows that many results from linear time invariant theory can be extended to periodic systems. A necessary condition for these results is that an equivalent time invariant system must be found for the periodic system. Misra provides a numerically reliable and simple procedure to find the equivalent time invariant systems, and he shows that a minimal-order generalized state-space description can always be found.

The research cited up to this point has focused on deterministic signal processing. In addition, the primary focus has centered on filter bank theory and how to best separate a signal into multiple parts that are sampled at different rates perform any necessary signal processing and then resynthesizing the processed signals.

In many signal processing techniques, however, applications involve signals that are random processes. Consequently, some work has been done to appropriately describe the statistical characterizations of multirate signal processing techniques.

Since multirate systems are inherently periodic in nature, signals within the multirate system are usually not wide-sense stationary (WSS). However, under appropriate conditions, stationary input signals can exhibit cyclostationary or wide-sense cyclostationary (WSCS) characteristics. In [Ref. 31] and [Ref. 32], Gardner presents methods on how to spectrally characterize N^{th} order cyclostationary signals and how to exploit spectral redundancies that might exist in cyclostationary signals. Related work in characterization of LTV systems was conducted by Akkarakaran and Vaidyanathan [Ref. 33] through the use of bifrequency and bispectrum maps. The bifrequency map is a two-dimensional Fourier transform used to characterize the LTV system while the bispectrum map is a two-dimensional Fourier transform that characterizes non-stationary random processes. In [Ref. 34] and [Ref. 35], Therrien presents methods for defining correlations functions and power spectra for multirate signals. This is extended by Therrien to non-stationary random processes in [Ref. 36] and in [Ref. 37].

Research into cyclic higher-order statistics of multirate signals has been conducted by Napolitano [Ref. 38]. Napolitano uses cyclic higher-order statistics to derive the input-output relations for MIMO linear almost-periodically time-variant systems excited by cyclostationary inputs. In addition, he presents a sufficient condition on the sampling rate to prevent aliasing when reconstructing cyclic higher-order statistics of continuous signals from sampled signals. Research into the problem of avoiding aliasing in the cyclic higher order spectra of a decimated time series is extended by Izzo and Napolitano in [Ref. 39] to a generalized form that avoids aliasing and imaging effects of a time series decimated by a fractional factor.

Using cyclostationary spectral analysis, Ohno and Sakai [Ref. 6] developed a method of deriving multirate optimal biorthogonal FIR filter banks that minimizes

the time-averaged mean-square error (TAMSE) after the high-frequency subband is removed. In order to properly describe the filter bank, cyclostationary spectral analysis is used since the output of the filter bank is cyclostationary for a wide-sense stationary input.

Other statistical methods are used by Yurdakul and Dundar [Ref. 40] to estimate the quantization error of FIR-based multirate systems. This is practical for implementation of multirate filters since the quantization of filter coefficients leads to quantization errors in the output signal of any discrete-time system.

One technique of statistical multirate signal processing researched by Sathe and Vaidayanathan [Ref. 41] is adaptive filtering. They use the statistical properties of signals in LTV systems to develop an adaptive filter structure that is useful for identification of band-limited channels. A matrix form of this adaptive filter is shown to provide better performance in terms of lower error energy than a traditional adaptive filter but suffers from high computational burden.

Outside of the signal processing literature, the most prominent technique in statistical multirate signal processing available is Kalman filtering. Many applications of and techniques for implementing multirate Kalman filters have been investigated. Early work on multirate Kalman filtering by Andrisani and Gau [Ref. 7] uses two different Kalman filters in parallel. One of the Kalman filters processes the high rate measurement, at a reduced order. The second Kalman filter processes the residuals of the first Kalman filter in conjunction with the low-rate measurement. Andrisani and Gau were able to reduce the computational complexity of their multirate Kalman filter by developing a suboptimal version, although a slight performance penalty was incurred.

Much work in Kalman filtering has been advanced by Bor-Sen Chen, along with You-Lin Chen and Chin-Wei Lin [Ref. 2, 3, 8, 9]. Through their work, the multirate Kalman filter has been applied to modeling of autoregressive (AR) and moving-average (MA) models through interpolation and estimation of the values of

the AR or MA stochastic signals. In addition, these researchers have developed methods for optimal signal reconstruction in noisy filter banks and developed a multirate Kalman filter that models system and channel noise. In addition to using a multirate Kalman reconstruction filter, they derived a sample-interpolation algorithm useful for the recovery of missing speech packets.

Further work on Kalman filtering was conducted by De Leon, Kober, Krumvieda and Thomas [Ref. 10]. By separating signals into subbands and then implementing multiple Kalman filters, a target can be tracked while reducing the computational rate and sensitivity to noisy measurements. Ni et al. [Ref. 11] also developed multirate Kalman filters that are less sensitive to noise. Their model combines the statistical model of the input signal with a multichannel representation of the subband signal. This form of the multirate Kalman filter provides the minimum variance reconstruction of the input signal, provided that the input signal is embedded in the state vector. Other applications of the multirate Kalman filter include orbit control of small aircraft [Ref. 13], helicopter passive ranging [Ref. 42, 43] and C.N.C. machining control [Ref. 12]. In addition, Tornero et al. [Ref. 14] combine the multirate Kalman filter with a multirate LQG controller to be used in self-location and path-tracking in mobile robots applications.

Another significant area of research employing multirate Kalman filtering is the multiresolution multirate (MRMR) estimation problem. One of the fundamental tools of MRMR techniques is the wavelet transform. The goal of MRMR estimation is to improve the overall estimation capability by exploiting features at different resolutions and sampling rates. Early work in this area was conducted by Basseville et al. [Ref. 44] on modeling and estimation of multiresolution statistical processes. Chou et al. [Ref. 45, 46] extended this work to recursive estimation and the kalman filter. More recently, Cristi and Tummala [Ref. 15] extended the work even further by developing a recursive MRMR Kalman filter.

In addition to multirate Kalman filtering, some research on multirate estima-

tion is available. Haddad, Bernstein and Huang developed reduced order estimators [Ref. 47]. These estimators were derived using a system of equations consisting of one modified Riccati equations and two modified Lyapunov equations, and each system of equation corresponds to a subinterval of the period associated with the multirate system. Another method of estimation was introduced by Hong [Ref. 48, 49]. His method focuses on using high-rate measurements to make low-rate estimations. By using filter banks and taking advantage of the lowpass filtering effect of the filter bank, improved approximations of the low rate estimation were obtained. In addition, a good approximation of the original measurements was obtained through orthogonal transformation by using highpass filtering and lowpass filtering. This thesis had its beginning in early work by Therrien and his students and there have been several publications along the way. In addition, some of the early work discussed here provided the groundwork for a companion thesis [Ref. 50] that extended the work along various other directions, most particularly to two-dimensional signals with applications to super-resolution image reconstruction. In [Ref. 51], Cristi, Koupatsiaris and Therrien present a method of multirate estimation for a two-signal input along with a quantitative analysis of reduction in mean-square error. The multirate estimator is extended to m-signals by Kuchler and Therrien in [Ref. 52] (preliminary work for this dissertation), and a recursive means of finding the filter coefficients and error variance as signals are added to the multirate system is presented. In addition, Therrien presents a method of linear prediction for multirate systems in [Ref. 53] along with a Levinson-type recursion for finding the prediction parameters (in conjunction with this dissertation). Further work by Therrien and Hawes in [Ref. 54] and [Ref. 55] present a method of least mean squares (LMS) calculation for multirate systems with two input signals. Lastly, some techniques used in the one-dimensional multirate system have been extended to the two-dimensional multirate system by Scrofani

and Therrien in [Ref. 56] and [Ref. 57]. They use multirate estimation algorithms to reconstruct a high-resolution image from a set of low resolution images that are subpixel translated (super-resolution).

B. DISSERTATION OUTLINE

From the literature review, it is apparent that the development of multirate filter banks has matured greatly. These filter banks, however, apply to SISO systems or MIMO systems where the sampling rates of all the input and output signals are the same. This dissertation begins by developing a general framework for multirate signal processing. The characterization of random processes as they are applied to decimators, expanders and filters are presented. Some LPTV filter bank realizations are discussed and, in particular, an LPTV filter bank that allows the input and output to be observed at different sampling rates is presented. Further simplification of signal presentation is presented through the use of matrix forms and Kronecker products. Whereas the majority of statistical multirate signal processing has focused on the Kalman filter, this dissertation approaches statistical multirate signal processing from the point of view of the more general multirate Wiener filter and its extensions.

The remainder of the dissertation is organized as follows. In Chapter II, the basic framework for describing a multirate system is developed. This includes describing the appropriate indexing schemes necessary to adequately describe different signals that are observed at different sampling rates and how they can be referenced to each other. This chapter also describes the different components available to multirate signal processing systems. In particular, the expander, decimator, linear time-invariant (LTI) filter and linear periodically time-varying (LPTV) filter are described. Further, the correlation equations for these components are presented. In Chapter III, the multirate Wiener-Hopf equations for optimal filtering are developed. The multirate optimal filter is presented both in its direct form and in its innovations

representation form. Simulations are conducted to analyze the performance of the multirate optimal filter. In Chapter IV, the multirate optimal filter is focused on the problem of linear prediction. A convenient formulation of this problem is presented and a novel efficient algorithm for solving the multirate linear prediction coefficients is developed. In Chapter V, the signal processing application of sequential classification is developed for a multirate system. Simulations are conducted to explore factors that affect the overall capabilities of multirate sequential classification. In addition, performance results for a single channel, a multichannel and a multirate system are compared. Finally, in Chapter VI, the research and conclusions of this dissertation are summarized and areas of further research are suggested.

THIS PAGE INTENTIONALLY LEFT BLANK

II. ANALYSIS OF MULTIRATE SIGNALS AND SYSTEMS

A. GENERAL CONCEPTS

Signal processing techniques have been well developed for processing a single-input and single-output system, or for processing a multiple-input and multiple-output system where all the signals are received at the same sampling rate; the latter is referred to as multichannel signal processing. In some signal processing situations, however, observed signals are sampled at various different rates, and it may be desired to process or exploit these signals together to perform various operations, such as estimation, prediction, detection or classification. To provide a basis for addressing some of these problems, some new theory needs to be developed that extends the theory for single-rate signals and systems. The theory developed for multirate statistical signal processing should reduce to the single-channel and single-rate multichannel problems as special cases. In this chapter, some basic concepts, notation, and terminology for describing stochastic multichannel systems and signals are introduced. In other words, this chapter presents the mathematical tools to be used in the remainder of the research.

A representation of a general multirate system is shown in Fig. 2.1. From Fig. 2.1 it can be seen that each signal has an associated sampling rate, F_i (Hz), where the letter i represents a particular signal. These sampling rates are generally different, although some may be identical. Since these signals, in general, have different sampling rates, their observations do not necessarily occur at the same point in continuous time and their indices m_i may not be aligned. For example, suppose a signal x is sampled at 3000 Hz and a signal y is sampled at 2000 Hz. If both signals have a sample that occurs at $t = 0$, then the next observation of the signal x would occur

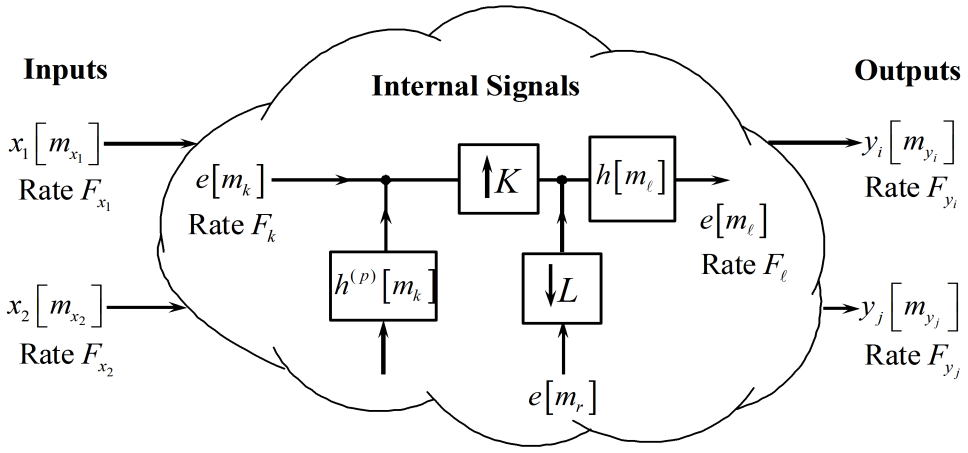


Figure 2.1. Multirate System Concept.

at 0.33 milliseconds and the next observation of y would occur at 0.5 milliseconds. Observations of both x and y would not coincide until 1 millisecond after the first observation of both signals. this concept is illustrated in Fig. 2.2. A diagram such as that provided in Fig. 2.2 will be called a sampling pattern.

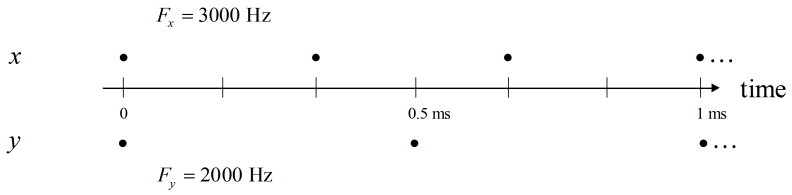


Figure 2.2. Effect of Different Sampling Rates on Observation Times.

In order to provide a common framework to which all the signals can be referenced, to account for the effects of sampling signals at different rates, a structure is defined here that will be called the *fundamental layer*. This structure does not necessarily correspond to any physical portion of the multirate system. It is a conceptual tool that is used to allow common referencing of all the signals within the multirate system.

A key element of the fundamental layer is that it has a sampling rate that is a multiple of the sampling rates for all signals in the system. This allows all of the signals to be described explicitly in terms of fundamental layer parameters. Returning to the example of two observed signals x and y with sampling rates $F_x = 3000$ Hz and $F_y = 2000$ Hz, respectively, the concept of the fundamental sampling rate is illustrated. In order to represent both signals in a common framework (the fundamental layer) for analysis, each signal is represented at a higher rate which is an integer multiple of the actual sampling rate. For efficiency, the lowest such rate that can be applied to all signals is used. In the example just described, this sampling rate would be 6000 Hz, and is illustrated by the tick marks on the time axis of Fig. 2.2. This sampling rate will be called the *fundamental rate*, $\overline{F}$ and its inverse $\overline{T} = 1/\overline{F}$ will be called the *system clock rate*. The fundamental rate is the minimum sampling rate necessary to describe all signals in the multirate system. It is assumed that each sampling rate F_i is integer-valued; thus the following definition can be made.

Definition 1. *The Fundamental Rate, $\overline{F}$, is given by*

$$\overline{F} = \mathrm{LCM}(F_1, F_2, \ldots), \tag{2.1}$$

where LCM( ) *denotes the* least common multiple *and F_1, F_2, ... represent the sampling rates of all the signals in the system.*

Each actual signal in the system can be associated with (a possibly fictitious) signal in the fundamental layer which is sampled at the fundamental rate $\overline{F}$. Thus each signal in the system can be thought of as a down-sampled or decimated version of some signal at the fundamental rate. The *decimation factor* associated with the i^{th} signal is given by

$$K_i = \frac{\overline{F}}{F_i}, \tag{2.2}$$

i.e., it is the ratio of the fundamental sampling rate to the signal sampling rate for the i^{th} signal of interest. While notation such as $x[m_x]$ is used to describe a signal at

some sampling rate F_x, the notation $\bar{x}[n]$ will be used to describe the corresponding signal in the fundamental layer. Thus these two signals are related by

$$x[m_x] = \bar{x}[n]\Big|_{n=K_x m_x} = \bar{x}[K_x m_x]$$

For a system conceptualized as in Fig. 2.1, a global periodicity exists. The *system period* represents the minimum number of time steps at the fundamental rate necessary to establish a repetitive sampling pattern for all signals in the multirate system.

Definition 2. *The System Period, K, is the minimum common digital period of all elements of the digital system. The system period is given by*

$$K = \mathrm{LCM}(K_1, K_2, \ldots), \tag{2.3}$$

where K_1, K_2, ... represent the decimation factors of all the signals in the system.

For the i^{th} signal, the number of *samples per period* is given by

$$M_i = \frac{K}{K_i}. \tag{2.4}$$

While K represents the number of time steps at the *system* rate corresponding to one period, M_i represents the number of time steps at the (i^{th}) *signal* rate corresponding to one period.

In order to demonstrate the concepts introduced thus far, the following example of a two-channel multirate system is provided. Consider a multirate system that has two signals x and y with sampling rates $F_x = 3000$ Hz and $F_y = 2000$ Hz, respectively. From (2.1), the fundamental rate for this system is $\overline{F} = \mathrm{LCM}(3000, 2000) = 6000$ Hz. Then, from (2.2), the signal decimation factors are calculated to be $K_x = 2$ and $K_y = 3$. Finally, from (2.3), the system period is found to be $K = 6$. Figure 2.3 shows the relationships among these terms. It can be seen from Fig. 2.3 that the decimation factors K_x and K_y represent the number of time steps, in the fundamental layer, between successive observations (samples) of a particular signal. Notice that

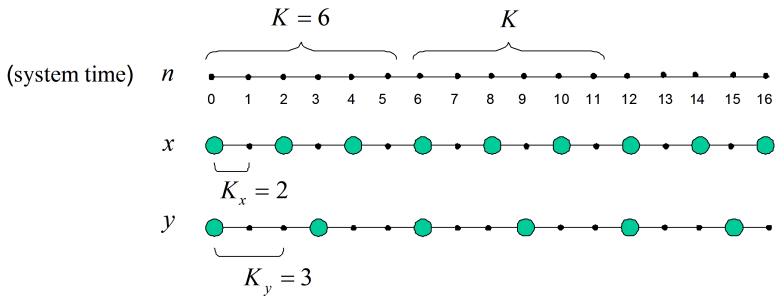


Figure 2.3. Illustration of Decimation Factors and System Period.

the sampling pattern repeats every $K = 6$ time steps at the fundamental rate. This also corresponds to $M_x = 3$ samples at the rate of x and $M_y = 2$ samples at the rate of y. Thus given a time n, the relative position within the sampling pattern is the same for all times $n + iK$ where i is an integer. This relative position will be called the *system phase* of the system (not to be confused with the phase of its Fourier transform). The system phase, denoted by k can be determined for any time n by writing

$$n = mK + k,$$

where $m = \lfloor n/K \rfloor$. Thus the following definition can be made.

Definition 3. *The System Phase, k, is given by*

$$n \equiv \ k \ (\text{mod } K), \tag{2.5}$$

where n is associated time at the fundamental layer, and k is the residual modulo K.

B. MULTIRATE SIGNAL PROCESSING

Some common forms of discrete-time signals can be directly related to an associated continuous-time signal. This will be shown using the sampling rates of the observed signals of a multirate system and the definition of the fundamental rate. Given a continuous time signal $x(t)$, the discrete-time signal for a signal sampled at

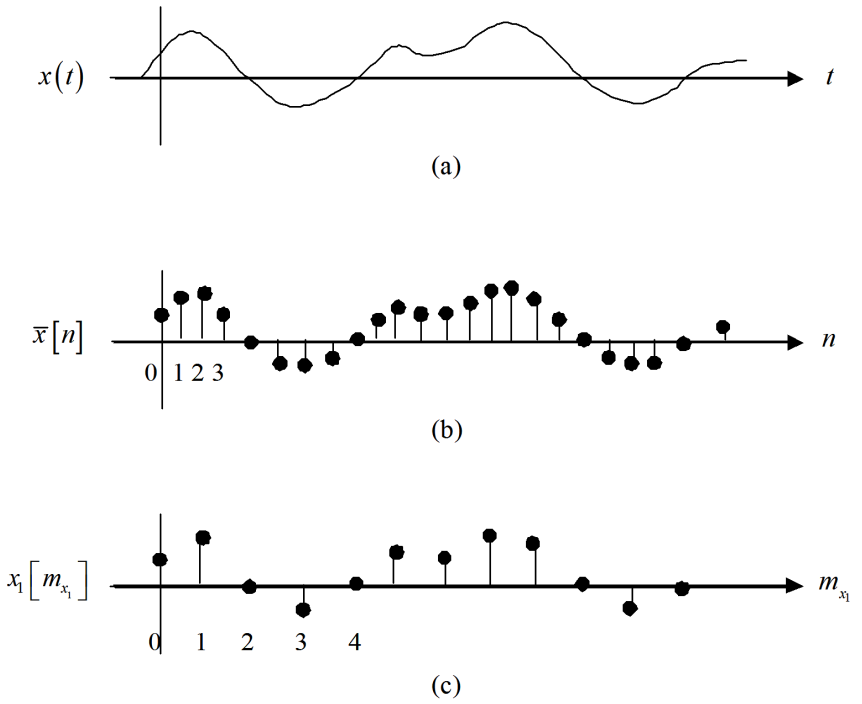


Figure 2.4. Continuous and Multirate Signals: (a) continuous signal, (b) discrete signal at fundamental sampling rate, (c) discrete signal at observed sampling rate.

the rate F_{x_i} would be $x_i[m_{x_i}] = x(m_{x_i}T_{x_i})$, where $T_{x_i} = \dfrac{1}{F_{x_i}}$ and $i = 1, \ldots, N$, where N is the number of signals in the system. Calculating the fundamental rate as $\overline{F} = \text{LCM}(F_{x_1}, F_{x_2}, \ldots, F_{x_N})$, the discrete-time signal associated with the fundamental rate would be $\bar{x}[n] = x(n\overline{T})$, where $\overline{T} = \dfrac{1}{\overline{F}}$ is the system clock rate. All signals are referenced to the initial time of $t = 0$, thus $x[0] = x(0T_x) = x(0)$. For clarity, parentheses are used to indicate the time dependence for a continuous signal, and brackets are used to indicate the time step for a discrete-time signal.

The relationships among some various signals are depicted in Fig. 2.4. The original continuous signal is denoted by $x(t)$. The signal in the fundamental layer is $\bar{x}[n]$, which may not exist in reality, and the signal which is sampled at the given rate F_{x_1} is then $x_1[m_{x_1}]$, which is a decimated version of $\overline{x}[n]$.

Some common multirate signal processing operations are now discussed in detail.

1. Decimation

The concept of decimation, or downsampling, was introduced earlier in Section II.A in conjunction with the definition of signals in the fundamental layer. In this section, decimation is discussed in a more general sense, as it may relate to various signals that may exist in a multirate system. Here decimation is discussed from the point of view of obtaining one signal within the system from another.

The process of decimation eliminates data points in a signal vector in order to reduce the sampling rate. The number of data points removed is dependent on the decimation factor. Given an integer-valued decimation factor, L, only one of every L consecutive samples is retained.

$$y[m_y] = x[Lm_y] \quad , \quad \text{for } m_y = \ldots, -1, 0, 1, \ldots . \tag{2.6}$$

The implied sampling rate for y is thus also reduced by a factor of L, i.e., $F_y = F_x/L$. The process of decimation is represented in block diagram form in Fig. 2.5, where L represents the decimation factor and m_y is the index of the decimated signal, such that m_y is the set of integers $Z = \{\ldots, -1, 0, 1, \ldots\}$. The associated index of the input signal is $m_x = Lm_y$.

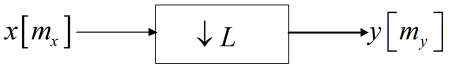


Figure 2.5. Decimation.

It is well known, [Ref. 17, 58], that the frequency spectrum $Y(\omega_y)$ of the decimated signal is related to the frequency spectrum $X(\omega_x)$ of the original signal by the formula

$$Y(\omega_y) = \frac{1}{L} \sum_{k=0}^{L-1} X\left(\frac{\omega_y}{L} - k\frac{2\pi}{L}\right). \tag{2.7}$$

The effect of decimation, in the frequency domain, is to stretch the frequency band as shown in Fig. 2.6. Notice that if the spectrum of the original signal is nonzero for

$|\omega| > \dfrac{\pi}{L}$, aliasing will occur due to undersampling. In most cases this is undesirable, although in some cases aliasing can be exploited to obtain useful results. To eliminate aliasing, the data sequence may be pre-filtered with an (ideal) low pass filter with a cutoff frequency at $|\omega| = \dfrac{\pi}{L}$.

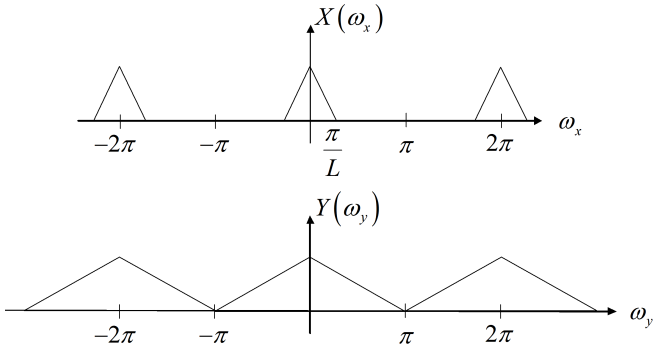


Figure 2.6. Frequency Spectrum for Decimation

Notice that given a signal $\overline{x}[n]$ at the fundamental rate, it is possible to generate L different decimated signals by applying a time shift to the signal at the fundamental rate before decimation. These L signals all have the same rate and are defined by

$$y^{(l)}[m] = \overline{x}[Lm_y + l],\ l = 0, 1, \ldots, L-1.$$

The generation of these signals is illustrated in Fig. 2.7,

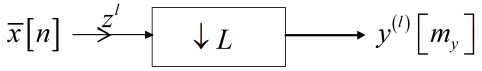


Figure 2.7. Decimation with Time Shift l.

where z^l represents a time shift of l units at the input rate. When all of the L possible decimation signals are formed, this process is known as *complete decimation*. Figure 2.8 illustrates complete decimation for a signal $\overline{x}[n]$ at the fundamental rate.

2. Expansion

The opposite of decimation is the process of expansion. Expansion, or up-sampling, increases the sampling rate by inserting zeros between the data points of a

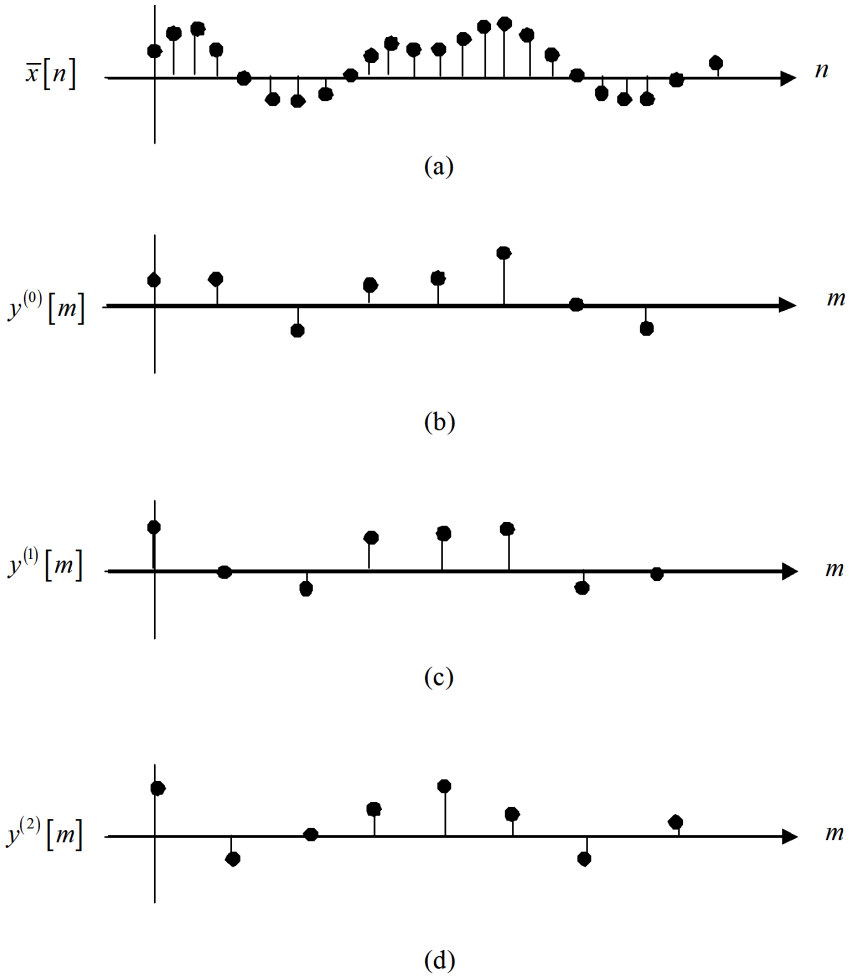


Figure 2.8. Complete decimation illustrated for $L = 3$: (a) discrete signal at fundamental sampling rate, (b) discrete signal at decimated sampling rate, no unit shift ($l = 0$), discrete signal at decimated sampling rate, one unit shift ($l = 1$), (d) discrete signal at decimated sampling rate, two unit shifts ($l = 2$).

signal. The sampling rate is increased by a factor of I if $I-1$ zeros are inserted after every data point. This process is defined by

$$y[m_y] = \begin{cases} x[m_y/I] & \text{for } m_y \text{ div } I \\ 0 & \text{otherwise} \end{cases} \tag{2.8}$$

where the expression "m_y div I" signifies that m_y is divisible by I, i.e., m_y/I is an integer. The process is represented in block diagram form in Fig. 2.9. The frequency

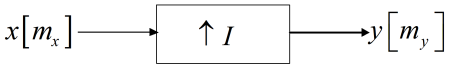


Figure 2.9. Expansion.

spectrum of the expanded signal is related to the original signal by

$$Y(\omega_y) = X(I\omega_y). \tag{2.9}$$

The effect of expansion in the frequency domain is to compress the frequency band as shown in Fig. 2.10. This compression brings additional copies of the original

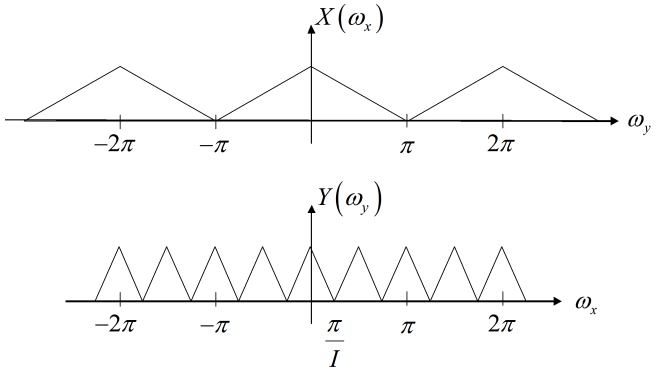


Figure 2.10. Frequency spectrum for Expansion.

spectrum, known as "images", into the frequency band of interest $(-\pi, \pi)$. In order to produce a rate change without distortion the expander must be followed by an (ideal) low-pass filter with cutoff frequency at $|\omega| = \dfrac{\pi}{I}$. In the time domain, filtering

after upsampling has the effect of "filling in" the zeros with "interpolated" values of the signal.

3. Rate Changes

Any rate change by a rational factor, I/L, can be achieved by expansion, low-pass filtering and then decimation of a signal. (It is assumed that I and L are coprime, i.e., these integers have no common factors.) The structure for this rate change system is shown in Fig. 2.11. The input $x[m_x]$ is at a rate F_x. After expansion, the rate changes to F_xI. The low-pass filter which is designed to operate at this new rate (F_xI) then serves both to eliminate the images due to expansion and provide bandlimiting so the decimated signal will not be aliased. In order to avoid aliasing, however, the cutoff frequency of the low pass filter must be less than or equal to $\frac{\pi}{C}$, where $C = \max(L, I)$. The output of the decimator is at the sampling rate $F_y = F_x \cdot I/L$.

Observe that the system rate for this simple system is $\overline{F} = F_xI$ and that the decimation factors are given by $K_x = I$ and $K_y = L$. Thus a rate change is accomplished by expanding up to the system rate, filtering and then decimating.

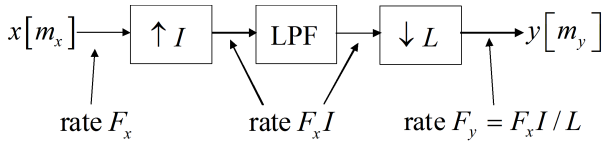


Figure 2.11. Rate Change by a Rational Factor I/L.

4. Filtering

Within the framework of multirate signal processing, there are two broad categories of filters that are most useful. These filters are 1) linear time-invariant filters and 2) linear periodically time-varying filters, and are discussed in the following subsubsections.

a. Linear Time-Invariant Filters

Linear time-invariant filters process a signal at some sampling rate F_x and produce an output at the same rate. The input x and output y are related by convolution as

$$y[m] = \sum_{k=-\infty}^{\infty} h[k]x[m-k] = \sum_{m=-\infty}^{\infty} x[m]h[n-m] \tag{2.10}$$

where $h[m]$ is called the impulse response and has the same rate as x and y. The system is *causal* if an only if $h[m] = 0$ for $m < 0$. In this case the lower limit of the variable m in (2.10) can be changed from $-\infty$ to 0.

The system is *stable* if and only if

$$\sum_{m=-\infty}^{\infty} \Big|h[m]\Big| < \infty.$$

The system is said to have finite impulse response (FIR) if $h[m]$ is a finite-length sequence and said to have infinite impulse response (IIR) otherwise. In the case of an FIR filter (2.10) can be reduced to a finite sum and the system is always stable.

b. Linear Periodically Time-Varying Filters

Linear periodically time-varying (LPTV) filters are important in multirate applications because of the inherent periodic nature of the multirate system. For example, in optimum filtering applications (see Chapter IV), the system periodicity requires that the filter coefficients change in a periodic fashion. This is illustrated in Fig. 2.12 where the filter output $y[m_y]$ is intended to estimate another signal $d[m_y]$.

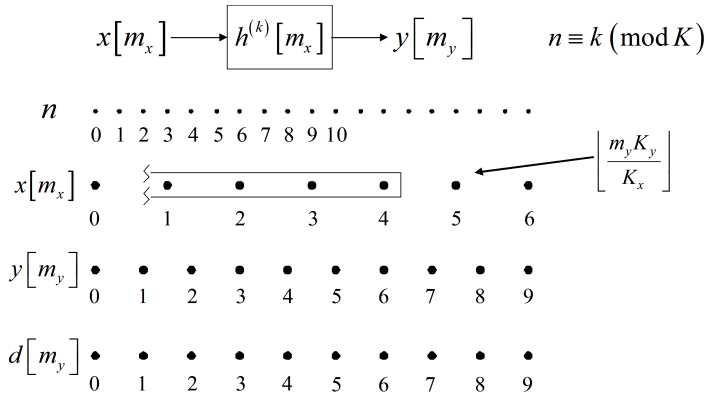


Figure 2.12. LPTV Filter

The sampling patterns for x, y and d are shown in Fig. 2.12 together with a time scale labeled 'n' that represents the system rate in the fundamental layer. The filtering is assumed to be *causal* so that at any selected time step only values of the input occurring at the same time or earlier on the system time scale are used. From Fig. 2.12 it can be seen that at time $m_y = 6$ the input sequence x and the output sequence y are aligned, however at time $m_y = 7$ the input sequence lags the output sequence. The difference in system clock time exhibited at these points implies there will be a difference in correlation between input and output. Therefore, to estimate the output properly at time m_y, a new set of filter coefficients different from those at time $m_y = 7$ are needed, although the filter coefficients are operating on the same input sequence.

Consider the following specific example. Given a causal LPTV filter with a filter order of four, the optimal estimate of d at time $m_y = 6$ is

$$\hat{y}[6] = h^{(0)}[0]x[4] + h^{(0)}[1]x[3] + h^{(0)}[2]x[2] + h^{(0)}[3]x[1], \tag{2.11}$$

where $h^{(0)}[l]$, $l = 0, 1, 2, 3$ are the optimal filter coefficients for estimating $d[6]$. At the next desired estimate time, $m_y = 7$, no new input data has been observed. Therefore, the estimate $\hat{y}[7]$ also is

$$\hat{y}[7] = h^{(2)}[0]x[4] + h^{(2)}[1]x[3] + h^{(2)}[2]x[2] + h^{(2)}[3]x[1], \tag{2.12}$$

but $h^{(2)}[l]$, $l = 0, 1, 2, 3$ is a different set of coefficients since the filter is estimating $d[7]$, instead of $d[6]$.

At time $m_y = 8$, a new observation of the input has already occurred. Therefore, the filter equation at this time instance is

$$\hat{y}[8] = h^{(4)}[0]x[5] + h^{(4)}[1]x[4] + h^{(4)}[2]x[3] + h^{(4)}[3]x[2]. \tag{2.13}$$

Again the coefficients are different from those occurring in (2.11) and (2.12). Finally, at time $m_y = 9$, the input observation occurs at the same time as the output estimate is calculated. This is the same system phase at the time instance of $m_y = 6$. Therefore, the optimal filter coefficients at time $m_y = 9$ are the same as those at time $m_y = 6$.

Although this example uses specific time steps to show how an LPTV filter functions, a generalized filter equation can be found. Let $\bar{x}$ and $\bar{y}$ be the representations of x and y in the fundamental layer, and let $\bar{h}^{(k)}[i]$ be a set of time-varying filter coefficients. The most generic linear filter can be written as

$$\bar{y}[n] = \sum_{i=-\infty}^{\infty} \bar{h}^{(k)}[i]\bar{x}[n-i]. \tag{2.14}$$

Observations of $\bar{x}$ only occur when its argument is a multiple of K_x however. Thus the right-hand side of (2.14) can be rewritten as

$$\bar{y}[n] = \sum_{m_x=-\infty}^{\infty} \bar{h}^{(k)}[K_x m_x]\bar{x}[K_x\lfloor n/K_x \rfloor - K_x m_x], \tag{2.15}$$

where $\lfloor\ \rfloor$ represents the floor (integer part) of its argument, and the term $K_x \cdot \lfloor n/K_x \rfloor$, on the right-hand side of the equation, is needed to account for fact that $\bar{x}$ is not observed at every time n. The most recent observation of x, in the fundamental layer, is described by $K_x\lfloor n/K_x \rfloor$. For example, if K_x is 3, then $\bar{x}$ is observed only at $n = 0, 3, 6, ...$ Thus, at $n = 5$, for instance, the most recent observation of x is at time $K_x\lfloor n/K_x \rfloor = 3\lfloor 5/3 \rfloor = 3 \cdot 1$ or $n = 3$. Now recall that x is a decimated version of $\bar{x}$ that satisfies

$$x[m_x] = \bar{x}[K_x m_x].$$

If a corresponding decimated version of the filter sequence is defined as

$$h^{(k)}[m] = \bar{h}^{(k)}[K_x m],$$

then the filter equation (2.15) can be expressed as

$$\bar{y}[n] = \sum_{m_x=-\infty}^{\infty} h^{(k)}[m_x] x[\lfloor n/K_x \rfloor - m_x]. \tag{2.16}$$

Further, observe that the output $\bar{y}[n]$ has values valid only at integer multiples of K_y, i.e., $n = m_y K_y$. Thus (2.16) becomes

$$\bar{y}[K_y m_y] = \sum_{m_x=-\infty}^{\infty} h^{(k)}[m_x] x[\lfloor K_y m_y/K_x \rfloor - m_x], \tag{2.17}$$

Finally, by noting that y is a decimated version of $\bar{y}$, one can write

$$\boxed{y[m_y] = \sum_{m_x=-\infty}^{\infty} h^{(k)}[m_x] x[\lfloor m_y K_y/K_x \rfloor - m_x], \qquad m_y \equiv\ k \pmod{K_x}.} \tag{2.18}$$

When the filtering is causal, the lower limit on the sum in (2.18) can be changed from $m_x = -\infty$ to $m_x = 0$ since the corresponding impulse response terms are zero.

Several implementations of LPTV filters using LTI filters are possible. Two common representations known in the literature are shown in Fig. 2.13 and Fig. 2.14. Figure 2.13 shows an LPTV filter realized in a rotary switch form. Given a system period K, the rotary switch steps through the sequence $0 \leq k \leq K-1$ in synchronization with the rate of the output sequence and selecting a different LTI filter to process the input. While this form is typically used in the literature to apply when the filter input and output rates are the same, it easily applies to the more general case represented by (2.18).

Figure 2.14 shows an LPTV filter realized in filter bank form. Here the output $y[n]$ is equal to the output of $H^{(k)}(z)$ where $n \equiv\ k \pmod{K}$. The purpose of decimation followed immediately by expansion is to remove any non-zero data not associated with the proper phase of the system (for a detailed explanation of this

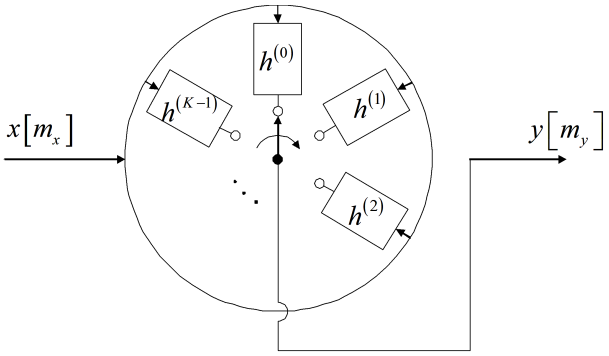


Figure 2.13. Rotary Switch Representation of an LPTV Filter

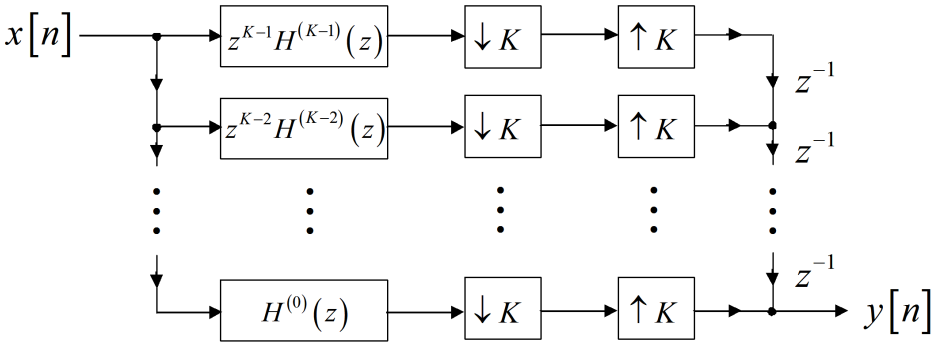


Figure 2.14. Filter Bank Representation of an LPTV Filter (From [Ref. 17])

filter bank see [Ref. 17]). The filter bank form of Fig. 2.14 is restricted to situations where the input and output are sampled at the same rate.

A generalized LPTV filter bank was derived that allows for the input $x[m_x]$ to be observed at a rate F_x and the output $y[m_y]$ to be observed at a rate F_y. This more general filter bank is depicted in Fig. 2.15. This filter bank operates at the fundamental rate $\overline{F}$, thus the input signal $x[m_x]$ needs to be expanded to the fundamental rate by K_x and the output needs to be decimated by K_y. In addition the delays are shifted by multiples of the output decimation factor K_y and there are M_y sets of filters $H^{(k_y)}(z)$, where M_y is the number of samples per period for signal y and $m_y \equiv k_y \pmod{M_y}$. If K_x and K_y are coprime, then $M_y = K_x$; but this is not true in general. Since $x[m_x]$ is observed at a rate $\dfrac{1}{K_x}$ of the clock rate, the filter

coefficients for the filters $H^{(k_y)}(z)$ are of the form

$$H^{(k_y)}(z) = h_0 + h_1 z^{-Kx} + h_2 z^{-2K_x} + \cdots + h_{P-1} z^{-(P-1)K_x},$$

where P is the filter order.

The LPTV filter bank of Fig. 2.15 requires a clock that operates at the fundamental rate $\overline{F}$. If two clocks are available that operate at the input and output rates, then the LPTV filter bank can be represented as shown in Fig. 2.16. In this form $z_y^{-1} = z^{-K_y}$ and $z_x^{-1} = z^{-K_x}$. The decimation and expansion factors are replaced with M_y since the right hand side of the filter is operating at the output rate F_y. In addition, the filters can be described by

$$H^{(k_y)}(z_x) = h_0 + h_1 z_x^{-1} + h_2 z_x^{-2} + \cdots + h_{P-1} z_x^{-(P-1)}.$$

C. STATISTICAL REPRESENTATION OF RANDOM SIGNALS

In order to adequately develop methods and models for processing random signals, it is important to be able to describe the statistical characteristics of the random signals. Since, in most cases the first and second moments are sufficient to describe the necessary statistical characteristics, this section is limited to "second

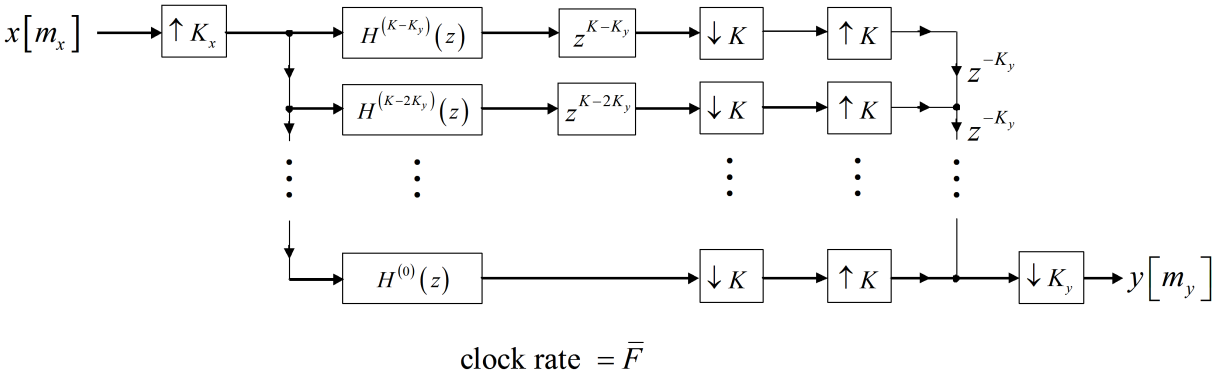


Figure 2.15. Filter Bank Representation of an LPTV Filter: Single Clock

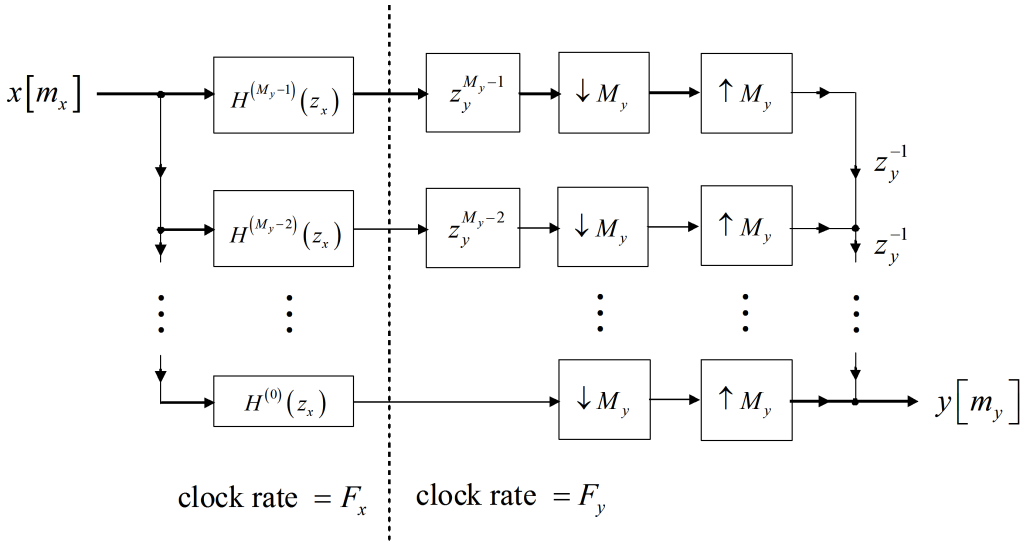


Figure 2.16. Filter Bank Representation of an LPTV Filter: Dual Clock

moment" properties. It will be seen that even with this limitation, there are significant issues to be addressed in the case of multirate signals and systems.

The first order moment, or *mean*, of a random process is defined as the expectation of the signal:

$$\mu_x[m] = \mathcal{E}\{x[m]\}. \tag{2.19}$$

The mean is, in general, a function of time, although for wide sense stationary (WSS) signals it is constant. For cases involving multirate signals and systems the mean is (in general) a *periodic* function of time, i.e., $\mu_x[m] = \mu_x[m + M]$ where M is the period.

The second order moments of a random process are defined via the autocorrelation function. Let m and m' represent two different values of the time index. Then the traditional *autocorrelation function* is defined as

$$R_x[m, m'] = \mathcal{E}\{x[m]x^*[m']\} \tag{2.20}$$

where '$*$' denotes the complex conjugate. The autocorrelation function is used for quantifying second order statistical linear dependence between samples of the signal at different points in time.

The *autocovariance* function is defined as the autocorrelation function with the mean removed:

$$\mathrm{C}_x[m, m'] = \mathcal{E}\{(x[m] - \mu_x[m])(x[m'] - \mu_x[m'])^*\}. \tag{2.21}$$

Frequently, the autocovariance is more useful than the autocorrelation, although the two functions are identical when the mean is zero. The autocorrelation and the autocovariance functions are related by,

$$R_x[m, m'] = \mathrm{C}_x[m, m'] + \mu_x[m]\mu_x^*[m']. \tag{2.22}$$

By defining the term *lag* as the difference between the two time instances, $l = m - m'$, the correlation and covariance functions can be represented also in terms of the time index, m, and the lag, l,

$$R_x[m; l] = \mathcal{E}\{x[m]x^*[m - l]\}. \tag{2.23}$$

This form is referred to as the "time-dependent" autocorrelation function, or the "time-lag" autocorrelation function. An analogous definition can be made for the autocovariance function.

Two special cases are important for multirate systems. First, if the random process is stationary, at least in the wide sense, then the mean of the random process is constant and the correlation and covariance functions are dependent on the lag only. Thus the dependence on m in (2.23) can be dropped and the autocorrelation function can be written as:

$$R_x[l] = \mathcal{E}\{x[m]x^*[m - l]\}.$$

In this case the traditional autocorrelation function $R_x[m, m']$, defined in (2.20) depends *only* on the difference $l = m - m'$.

A second case of importance for multirate systems is the cyclostationary case. A more complete discussion of cyclostationarity can be found in [Ref. 31, 32], however,

it is simply observed here that for a cyclostationary process the time-lag autocorrelation function is a periodic function of m, i.e.,

$$R_x[m;l] = R_x[m+M;l], \quad \exists M \tag{2.24}$$

where M is the period. This equation will be taken here as the *definition* of a cyclostationary process. Note that $R_x[m;l]$ is not necessarily periodic in l. Although for a cyclostationary random process the traditional autocorrelation function $R_x[m,m']$ is periodic in *both* arguments, with the *same* period M (See [Ref. 31, 32].)

For two signals sampled at the *same* rate the traditional cross-correlation function is defined as

$$R_{xy}[m,m'] = \mathcal{E}\{x[m]y^*[m']\},$$

where m and m' are two arbitrary values of the index. The cross-correlation can also be represented in the time-lag form as

$$R_{xy}[m;l] = \mathcal{E}\{x[m]y^*[m-l]\},$$

where $l = m - m'$. Further if the signals are jointly wide-sense stationary, then the cross-correlation is a function of its lag term only,

$$R_{xy}[l] = \mathcal{E}\{x[m]y^*[m-l]\}.$$

In multirate signal processing however, one must be able to find the cross-correlation of signals at different rates. This requires more careful consideration. The discussion here follows the development in [Ref. 35].

The cross-correlation function for signals in the fundamental layer can be defined in two forms:

$$\bar{R}_{\bar{x}\bar{y}}[n_x,n_y] = \mathcal{E}\{\bar{x}[n_x]\bar{y}^*[n_y]\} \tag{2.25}$$

and

$$\bar{R}_{\bar{x}\bar{y}}[n;\nu] = \mathcal{E}\{\bar{x}[n]\bar{y}^*[n-\nu]\}. \tag{2.26}$$

The latter is the time-lag form.

The correlation function for two signals at different rates can be defined in the traditional form. Following the discussion in Section II.B, $x[m_x]$ and $y[m_y]$ are represented in terms of signals in the fundamental layer. This representation makes it straightforward to write

$$\begin{aligned} R_{xy}[m_x, m_y] &= \mathcal{E}\{x[m_x]y^*[m_y]\} = \mathcal{E}\{\overline{x}[K_x m_x]y^*[K_y m_y]\} \\ &= \bar{R}_{\bar{x}\bar{y}}\left[K_x m_x, K_y m_y\right]. \end{aligned} \tag{2.27}$$

One can consider the cross-correlation function $R_{xy}[m_x, m_y]$ as samples of the cross-correlation $\bar{R}_{\bar{x}\bar{y}}$ in the fundamental layer, as shown in Fig. 2.17. The samples are defined on a rectangular grid or *lattice* [Ref. 59, 60] of points separated by K_x samples of the fundamental layer in the x direction and K_y samples in the y direction (see Fig. 2.17). Correlation for values of n_x and n_y not on this lattice are not defined.

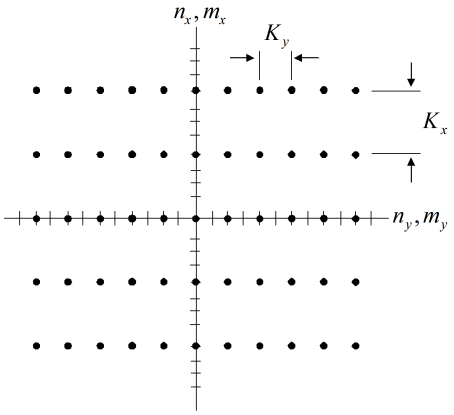


Figure 2.17. Cross-correlation Lattice.

Developing the time-lag form of correlation for two signals at different rates is somewhat more complex. Consider the following definition, which can be written by analogy:

$$R_{xy}[m; l] = \mathcal{E}\{x[m]y^*[m-l]\} = \mathcal{E}\{\overline{x}[K_x m]\bar{y}^*[K_y(m-l)]\}, \quad \begin{matrix} m = \ldots, -1, 0, 1, \ldots \\ l = \ldots, -1, 0, 1, \ldots \end{matrix} \tag{2.28}$$

With ν as the time delay in the fundamental layer, the cross-correlation can be plotted in the (n, ν) space as shown in Fig. 2.18. The indices m and l map to points on a non-rectangular lattice. Note that correlation values do not exist for all possible combinations of n and ν because samples of $\bar{y}$ occur only when its argument is an integer multiple of K_y.

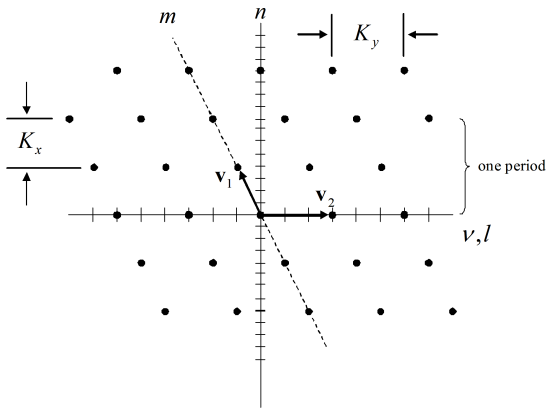


Figure 2.18. Time-lag Cross-correlation Lattice.

The lattice is described by two basis vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ which are defined in the fundamental layer as

$$\mathbf{v}_1 = \begin{bmatrix} K_x \\ K_x - K_y \end{bmatrix} \quad \text{and} \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ K_y \end{bmatrix}$$

The indices m and l in $R_{xy}[m; l]$ represent a point on the lattice that is reached by taking m steps in the $\mathbf{v}_1$ direction and l steps in the $\mathbf{v}_2$ direction. The time-lag correlation function and the traditional cross-correlation function are related as

$$R_{xy}[m; l] = R_{xy}[m, m - l]. \tag{2.29}$$

When the signals $\bar{x}$ and $\bar{y}$ are jointly wide sense stationary, then the cross-correlation as defined in (2.28) is a function of ν only (and not n). Therefore, the values of correlation depend only on the distance from the n axis in Fig. 2.18 and since the sampling pattern is periodic, the values of correlation repeat periodically.

Consider the following example. Suppose the cross-correlation function for signals in the fundamental layer is given by

$$\bar{R}_{\bar{x}\bar{y}}[n_x, n_y] = \beta^{-|n_x - n_y|}, \qquad (|\beta| < 1), \tag{2.30}$$

where n_x represents the time instance for the signal x in the fundamental layer and n_y represents the time instance of y in the fundamental layer. Note that the symbol $\bar{R}_{\bar{x}\bar{y}}$ is used to distinguish the cross-correlation of the signals in the fundamental layer from R_{xy}, the cross-correlation of the associated decimated signals.

From (2.27), the cross-correlation of the decimated signals is

$$\begin{aligned} R_{xy}[m_x, m_y] &= \mathcal{E}\{\bar{x}[K_x m_x]\bar{y}^*[K_y m_y]\} \\ &= \bar{R}_{xy}[K_x m_x, K_y m_y] \\ &= \beta^{-|K_x m_x - K_y m_y|} \end{aligned}$$

Then from (2.29), the cross-correlation can be represented in the (m, l) space as

$$\begin{aligned} R_{xy}[m; l] &= R_{xy}[m, m - l] \\ &= \beta^{-|K_x m - K_y (m-l)|} \\ &= \beta^{-|(K_x - K_y)m + K_y l|}. \end{aligned} \tag{2.31}$$

Note that the cross-correlation of the multirate signals is a subset of the values of cross-correlation of the signals in the fundamental layer.

In order to further illustrate the relation between the (m_x, m_y) and (m, l) spaces, the following example is provided. In this example numerical values are used to show how valid data points in the (m_x, m_y) space are transformed into valid data points in the (m, l) space. Assuming $K_x = 2$ and $K_y = 3$ and using the definitions that $m = m_x$ and $l = m_x - m_y$, Table 2.1 shows the results of the transformations and Fig. 2.19 visually depicts how the points map from one space to the other.

Table 2.1. Index transformation of the cross-correlation

(m_x, m_y)	$(0,1)$	$(1,0)$	$(1,2)$	$(2,1)$	$(2,2)$
$(m;l)$	$(0;-1)$	$(1;1)$	$(1;-1)$	$(2;1)$	$(2;0)$
Symbol	◯	□	△	◊	▽

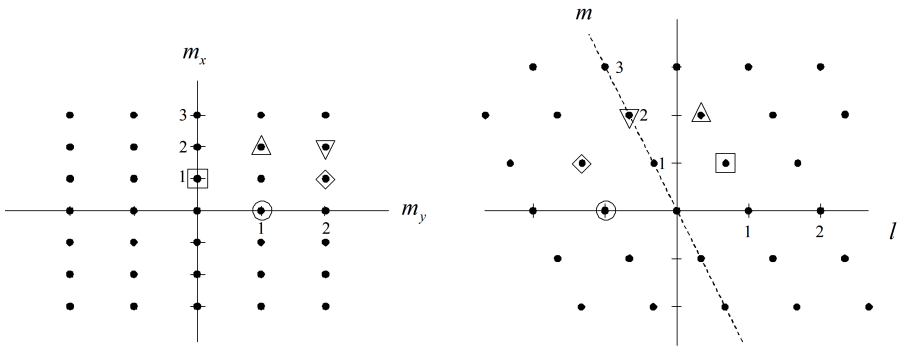


Figure 2.19. Index Transformation of the Cross-correlation.

D. INPUT-OUTPUT CROSS-CORRELATION FOR EXPANSION, DECIMATION AND FILTERING

Having defined the mean, autocorrelation and cross-correlation functions, expressions can be derived for the relationships between these statistics for the input and output for some common operations. In particular, the operations of decimation, expansion and linear filtering are considered. In the following, it is assumed that the autocorrelation function $R_x[m_x, m'_x]$ of the input as given by (2.20) is known. Equivalently, the time-lag form of the autocorrelation $R_x[m_x; l_x]$ of the input as defined by (2.23) is known. Further, if the signal is wide-sense stationary, the time-lag autocorrelation function can be reduced to

$$R_x[l_x] = \mathcal{E}\{x[m_x]x^*[m_x - l_x]\}. \tag{2.32}$$

1. Decimation

The operation of decimation is depicted in Fig. 2.5 and defined as (2.6). For ease of reference, the mathematical description of decimation is also provided below:

$$y[m_y] = x[Lm_y] \quad \text{for } m = \ldots, -1, 0, 1, \ldots.$$

Given two arbitrary time instances, m_y and m'_y, the output autocorrelation for the decimator is given by

$$\begin{aligned} R_y[m_y, m'_y] &= \mathcal{E}\{y[m_y]y^*[m'_y]\} \\ &= \mathcal{E}\{x[Lm_y]x^*[Lm'_y]\} \\ R_y[m_y, m'_y] &= R_x[Lm_y, Lm'_y] \end{aligned} \tag{2.33}$$

Thus decimation results in decimation of the auto-correlation function.

By defining a lag term which is the difference between the two time instances, $l = m_y - m'_y$ and renaming, $m = m_y$, the autocorrelation function can be represented in the time-lag form as

$$\begin{aligned} R_y[m; l] &= \mathcal{E}\{y[m]y^*[m - l]\} \\ &= \mathcal{E}\{x[Lm]x^*[Lm - Ll]\} \end{aligned}$$

or

$$R_y[m; l] = R_x[Lm; Ll] \tag{2.34}$$

Further, if the signal is wide-sense stationary, then the autocorrelation function is independent of m and can be written as

$$R_y[l] = R_x[Ll] \tag{2.35}$$

The *cross*-correlation function for decimation is given by

$$\begin{aligned} R_{xy}[m_x, m_y] &= \mathcal{E}\{x[m_x]y^*[m_y]\} \\ &= \mathcal{E}\{x[m_x]x^*[Lm_y]\} \end{aligned}$$

or

$$R_{xy}[m_x, m_y] = R_x[m_x, Lm_y] \tag{2.36}$$

Thus the original cross-correlation function is decimated in one argument only. Figure 2.20 visually depicts the effect of decimation on the cross-correlation function. The circled locations of the input autocorrelation R_x are the values preserved in the cross-correlation of the decimator.

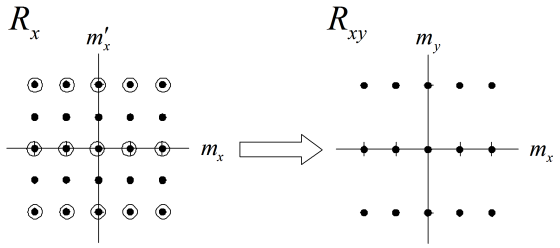


Figure 2.20. Relation of the Input Autocorrelation to the Cross-correlation of the Decimator.

Using the time-lag definition for the cross-correlation of two signals at different sampling rates, the cross-correlation of the decimator can be written as

$$\begin{aligned} R_{xy}[m_x; l_y] &= \mathcal{E}\{x[m_x]y^*[m_x - l_y]\} \\ &= \mathcal{E}\{x[m_x]x^*[L(m_x - l_y)]\} \\ &= \mathcal{E}\{x[m_x]x^*[m_x - (Ll_y - (L-1)m_x)]\} \end{aligned}$$

or

$$R_{xy}[m_x; l_y] = R_x[m_x; Ll_y - (L-1)m_x] \tag{2.37}$$

The expression for the time-lag cross-correlation function in terms of the auto-correlation function is rather cumbersome. The geometric representation in the (m, ν) is more clear. The effect of decimation on the cross-correlation in the $(m; l)$ space is visually depicted in Fig. 2.21. Again, the circled values of the input autocorrelation are preserved in the cross-correlation.

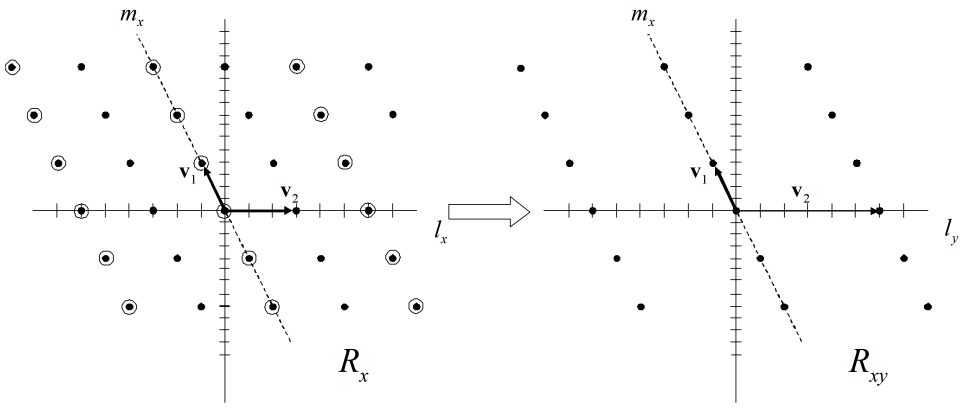


Figure 2.21. Relation of the Input Autocorrelation to the Cross-correlation of the Decimator in the $(m; l)$ Space.

Even when the input is wide-sense stationary, the cross-correlation cannot be described by its lag term only, since the lag term on the right-hand side of (2.37) requires knowledge of m_x.

2. Expansion

Again, for ease of reference, the mathematical description of expansion is provided below:

$$y[m] = \begin{cases} x[m/I] & m \text{ div } I \\ 0 & \text{otherwise} \end{cases}$$

The autocorrelation of the output signal is defined as

$$\begin{aligned} R_y[m_y, m'_y] &= \mathcal{E}\{y[m_y]y^*[m'_y]\} \\ &= \begin{cases} \mathcal{E}\{x[m_y/I]x^*[m'_y/I]\} & m_y, m'_y \text{ div } I \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Therefore,

$$R_y[m_y, m'_y] = \begin{cases} R_x[m_y/I], m'_y/I] & m_y, m'_y \text{ div } I \\ 0 & \text{otherwise} \end{cases}$$

where m_y and m'_y represent two arbitrary values of the time index.

The time-lag form of the autocorrelation can be written as

$$R_y[m_y; l_y] = \mathcal{E}\{y[m_y]y^*[m_y - l_y]\}$$
$$= \begin{cases} \mathcal{E}\{x[m_y/I]x^*[(m_y - l_y)/I]\} & m_y, (m_y - l_y) \text{ div } I \\ 0 & \text{otherwise} \end{cases}$$

Therefore,

$$R_y[m_y; l_y] = \begin{cases} R_x[m_y/I]; l_y/I] & m_y, l_y \text{ div } I \\ 0 & \text{otherwise} \end{cases}$$

In the case of the expander, the output is never wide-sense stationary even if the input is WSS, because zeros have been added. Therefore, the autocorrelation of the expander cannot be represented by a lag only, and a time reference must be stated as well. If the input is WSS, the output is cyclostationary however, because the autocorrelation function is periodic in its first argument.

The cross-correlation for expansion is computed as

$$R_{xy}[m_x, m_y] = \mathcal{E}\{x[m_x]y^*[m_y]\}$$
$$= \begin{cases} \mathcal{E}\{x[m_x]x^*[m_y/I]\} & m_y \text{ div } I \\ 0 & \text{otherwise} \end{cases}$$

Therefore,

$$R_{xy}[m_x, m_y] = \begin{cases} R_x[m_x, m_y/I] & m_y \text{ div } I \\ 0 & \text{otherwise} \end{cases}$$

Thus the original cross-correlation function is expanded in one argument only. Figure 2.22 visually depicts the effect of expansion on the cross-correlation function. The circled locations of the expander cross-correlation are the autocorrelation values preserved from the input. All other values are zero.

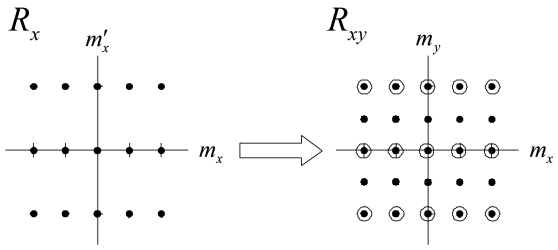


Figure 2.22. Relation of the Input Autocorrelation to the Cross-correlation of the Expander.

Using the time-lag definition for the cross-correlation of two signals at different sampling rates, the cross-correlation of the expander can be written as

$$\begin{aligned} R_{xy}[m_x; l_y] &= \mathcal{E}\{x[m_x]y^*[m_x - l_y]\} \\ &= \begin{cases} \mathcal{E}\{x[m_x]x^*[(m_x - l_y)/I]\} & (m_x - l_y) \text{ div } I \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} \mathcal{E}\{x[m_x]x^*[m_x - (l_y + (I-1)m_x)/I)]\} & (m_x - l_y) \text{ div } I \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Therefore,

$$R_{xy}[m_x; l_y] = \begin{cases} R_x[m_x; (l_y + (I-1)m_x)/I)]\} & m_x - l_y \text{ div } I \\ 0 & \text{otherwise} \end{cases} \tag{2.38}$$

The effect of expansion on the cross-correlation in the (m, l) space is visually depicted in Fig. 2.23. Again, the circled values of the cross-correlation are the values preserved from the input autocorrelation. The uncircled points represent valid locations where the cross-correlation is defined, but these values are zero.

Similar to the decimator, when the input of the expander is wide-sense stationary the output cannot be represented in terms of its lag only. The first reason is that the lag term on the right-hand side of (2.38) requires knowledge of m_x. In

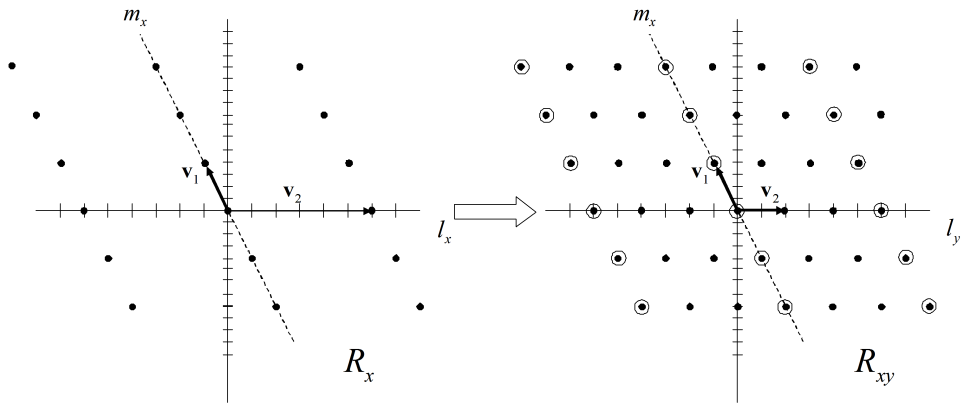


Figure 2.23. Relation of the Input Autocorrelation to the Cross-correlation of the Expander in the (m, l) Space.

addition, the expander inserts zeros, therefore the output is not WSS. It is however, wide-sense cyclostationary.

3. Linear Time-invariant Filters

For the operation of linear *time-invariant* filtering the output is given by

$$y[m] = \sum_{r=-\infty}^{\infty} h[r]x[m-r].$$

Given two arbitrary time instances, m and m', the autocorrelation of the filter output is

$$\begin{aligned} R_y[m,m'] &= \mathcal{E}\{y[m]y^*[m']\} \\ &= \mathcal{E}\left\{\left(\sum_{R_1=-\infty}^{\infty} h[R_1]x[m-R_1]\right)\left(\sum_{R_0=-\infty}^{\infty} h[R_0]x[m'-R_0]\right)^*\right\} \\ &= \mathcal{E}\left\{\sum_{R_1=-\infty}^{\infty}\sum_{R_0=-\infty}^{\infty} h[R_1]h^*[R_0]x[m-R_1]x^*[m'-R_0]\right\} \\ &= \sum_{R_1=-\infty}^{\infty}\sum_{R_0=-\infty}^{\infty} h[R_1]h^*[R_0]R_x[m-R_1,m'-R_0] \end{aligned}$$

This can be written as a two-dimensional convolution,

$$R_y[m,m'] = R_x[m,m'] * h[m] * h^*[m']$$

where '$*$' represents convolution with respect to the associated variable.

By using the relation $R_x[m;l] = R_x[m, m-l]$, where $l = m - m'$, the time-lag form of the output autocorrelation becomes

$$R_y[m;l] = \sum_{R_1=-\infty}^{\infty} \sum_{R_0=-\infty}^{\infty} h[R_1]h^*[R_0]R_x[m - R_1; l - (R_1 - R_0)]$$

For a stationary process, the time-lag correlation function is independent of the first argument. Therefore, the result for stationary processes becomes

$$R_y[l] = \sum_{R_1=-\infty}^{\infty} \sum_{R_0=-\infty}^{\infty} h[R_1]h^*[R_0]R_x[l - (R_1 - R_0)].$$

This can be put in the well-known form

$$R_y[l] = R_x[l] * h[l] * h^*[-l],$$

where '$*$' represents convolution.

The traditional cross-correlation function between the input and the output of the filter is given by

$$\begin{aligned} R_{xy}[m, m'] &= \mathcal{E}\{x[m]y^*[m']\} \\ &= \mathcal{E}\left\{x[m]\left(\sum_{R_0=-\infty}^{\infty} h[R_0]x[m' - R_0]\right)^*\right\} \\ &= \mathcal{E}\left\{\sum_{R_0=-\infty}^{\infty} h^*[R_0]x[m]x^*[m' - R_0]\right\} \\ &= \sum_{R_0=-\infty}^{\infty} h^*[R_0]R_x[m, m' - R_0] \end{aligned}$$

or

$$R_{xy}[m, m'] = R_x[m, m'] * h^*[m']$$

Again using the relation $R_{xy}[m;l] = R_{xy}[m, m-l]$, where $l = m - m'$, the time-lag form of the output cross-correlation function is:

$$R_{xy}[m;l] = \sum_{R_0=-\infty}^{\infty} h^*[R_0]R_x[m; l + R_0]$$

or

$$R_{xy}[m;l] = R_x[m;l] * h^*[-l].$$

When the input process is stationary, this reduces to

$$R_{xy}[l] = R_x[l] * h^*[-l].$$

4. Linear Periodically Time-varying Filters

When calculating the correlation functions for a LPTV filter, care must be taken to account for the difference in rates between the input and output. Recall that the general equation for an LPTV filter is given by (2.18).

The traditional cross-correlation function between the input and the output of the LPTV filter is thus

$$\begin{aligned} R_{xy}[m_x, m_y] &= \mathcal{E}\{x[m_x]y^*[m_y]\} \\ &= \mathcal{E}\left\{x[m_x]\left(\sum_{R_0=-\infty}^{\infty} h^{(k)}[R_0]x[\lfloor m_y K_y/K_x\rfloor - R_0]\right)^*\right\} \\ &= \mathcal{E}\left\{\sum_{R_0=-\infty}^{\infty} h^{*(k)}[R_0]x[m_x]x^*[\lfloor m_y K_y/K_x\rfloor - R_0]\right\} \\ &= \sum_{R_0=-\infty}^{\infty} h^{*(k)}[R_0]R_x[m_x, \lfloor m_y K_y/K_x\rfloor - R_0]. \end{aligned}$$

which represents a convolution for $\mathbf{h}^{*(k)}$ with R_x along its first argument.

The autocorrelation for the LPTV filter is given by

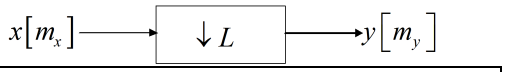
$$\begin{aligned} R_y[m_y, m_y'] &= \mathcal{E}\{y[m_y]y^*[m_y']\} \\ &= \mathcal{E}\left\{\left(\sum_{R_1=-\infty}^{\infty} h^{(k)}[R_1]x[\lfloor m_y K_y/K_x\rfloor - R_1]\right)\left(\sum_{R_0=-\infty}^{\infty} h^{(k')}[R_0]x[\lfloor m_y' K_y/K_x\rfloor - R_0]\right)^*\right\} \\ &= \mathcal{E}\left\{\sum_{R_1=-\infty}^{\infty}\sum_{R_0=-\infty}^{\infty} h^{(k)}[R_1]h^{*(k')}[R_0]x[\lfloor m_y K_y/K_x\rfloor - R_1]x^*[\lfloor m_y' K_y/K_x\rfloor - R_0]\right\} \\ &= \sum_{R_1=-\infty}^{\infty}\sum_{R_0=-\infty}^{\infty} h^{(k)}[R_1]h^{*(k')}[R_0]R_x[\lfloor m_y K_y/K_x\rfloor - R_1, \lfloor m_y' K_y/K_x\rfloor - R_0], \end{aligned}$$

where $m_y \equiv k \pmod{K_x}$ and $m'_y \equiv k' \pmod{K_x}$. This equation is the two-dimensional convolution along indices m_y and $m_y^{'}$.

Because of the floor operation used to transform between the m_y index and the m_x index, the time-lag equivalents for the autocorrelation and the cross-correlation functions become rather unwieldly.

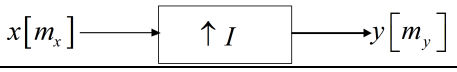
A summary of correlation relations is provided in Tables 2.2 through 2.4. A summary of relations in the frequency domain is provided in Appendix A. These relations were not used in this dissertation, so are not presented here, but are provided in the appendix for reference.

Table 2.2. Summary of Decimation



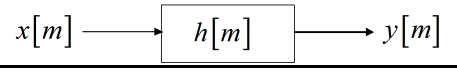
$R_y[m_y, m_y^{'}] = R_x[Lm_y, Lm_y^{'}]$
$R_y[m_y; l_y] = R_x[Lm_y; Ll_y]$
$R_y[l_y] = R_x[Ll_y]$
$R_{xy}[m_x, m_y] = R_x[m_x, Lm_y]$
$R_{xy}[m_x; l_y] = R_x[m_x; Ll_y - (L-1)m_x]$

Table 2.3. Summary of Expansion



$R_y[m_y, m'_y] = \begin{cases} R_x[m_y/I, m'_y/I] & m_y, m'_y \text{ div } I \\ 0 & \text{otherwise} \end{cases}$
$R_y[m_y; l_y] = \begin{cases} R_x[m_y/I; l_y/I] & m_y, l_y \text{ div } I \\ 0 & \text{otherwise} \end{cases}$
$R_{xy}[m_x, m_y] = \begin{cases} R_x[m_x, m_y/I] & m_y \text{ div } I \\ 0 & \text{otherwise} \end{cases}$
$R_{xy}[m_x; l_y] = \begin{cases} R_x[m_x; (l_y + (I-1)m_x)/I] & m_x - l_y \text{ div } I \\ 0 & \text{otherwise} \end{cases}$

Table 2.4. Summary of Filtering



$R_y[m, m'] = R_x[m, m'] * h[m] * h^*[m']$
$R_y[l] = R_y[l] * h[l] * h^*[-l]$
$R_{xy}[m, m'] = R_x[m, m'] * h^*[m']$
$R_{xy}[m, l] = R_x[m; l] * h^*[-l]$

E. MATRIX REPRESENTATION

In order to develop signal processing algorithms in the time domain, it is frequently easier to work with matrix and vector representations of the systems and signals. This section therefore develops tools to represent the basic operations of decimation, expansion and linear filtering. The development extends the ideas described in [Ref. 61].

1. Decimation

Consider the decimator depicted in Fig. 2.5. The input and output are related by $y[m] = x[Lm]$. Define the two vectors of input and output samples

$$\mathbf{x} = \begin{bmatrix} x[0] & x[1] & \cdots & x[LM-1] \end{bmatrix}^T$$

$$\mathbf{y} = \begin{bmatrix} y[0] & y[1] & \cdots & y[M-1] \end{bmatrix}^T.$$

Then the decimation operation can be represented as

$$\mathbf{y} = \mathbf{D}_{L,M}\mathbf{x},$$

where $\mathbf{D}_{L,M}$ is an $M \times ML$ matrix of 1's and 0's used to extract elements of $\mathbf{x}$ as observations to be placed in $\mathbf{y}$. The matrix $\mathbf{D}_{L,M}$ will be called a *decimation matrix*. For example, if the decimation factor is $L = 3$ and the size of the desired output vector is $M = 4$, then the associated decimation matrix is

$$\mathbf{D}_{3,4} = \begin{bmatrix} 1\,0\,0 & 0\,0\,0 & 0\,0\,0 & 0\,0\,0 \\ 0\,0\,0 & 1\,0\,0 & 0\,0\,0 & 0\,0\,0 \\ 0\,0\,0 & 0\,0\,0 & 1\,0\,0 & 0\,0\,0 \\ 0\,0\,0 & 0\,0\,0 & 0\,0\,0 & 1\,0\,0 \end{bmatrix}. \tag{2.39}$$

This can be represented in a more compact form using Kronecker products. In this compact form

$$\mathbf{D}_{L,M} = \mathbf{I}_M \otimes \mathbf{d}_L^T,$$

where $\mathbf{I}_M$ is the $M \times M$ identity matrix and $\mathbf{d}_L$ is a column vector consisting of a one followed by $L-1$ zeros,

$$\mathbf{d}_L = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \end{bmatrix}^T. \tag{2.40}$$

Appendix B provides a short discussion of Kronecker products. For a detailed analysis of Kronecker products and matrix calculus see [Ref. 62], or for a short tutorial paper see [Ref. 63].

Using this definition and dropping the subscripts for ease of notation, the mean of the decimator is

$$\mathbf{m}_y = \mathcal{E}\{\mathbf{y}\} = \mathcal{E}\{\mathbf{D}\mathbf{x}\} = \mathbf{D}\mathcal{E}\{\mathbf{x}\} = \mathbf{D}\mathbf{m}_x, \tag{2.41}$$

and the output autocorrelation matrix for the decimator is

$$\mathbf{R}_y = \mathcal{E}\{\mathbf{y}\mathbf{y}^{*T}\} = \mathcal{E}\{\mathbf{D}\mathbf{x}(\mathbf{D}\mathbf{x})^{*T}\} = \mathbf{D}\mathcal{E}\{\mathbf{x}\mathbf{x}^{*T}\}\mathbf{D}^{*T} = \mathbf{D}\mathbf{R}_x\mathbf{D}^{*T}. \tag{2.42}$$

Likewise, the matrix form of the cross-correlation is

$$\mathbf{R}_{xy} = \mathcal{E}\{\mathbf{x}\mathbf{y}^{*T}\} = \mathcal{E}\{\mathbf{x}(\mathbf{D}\mathbf{x})^{*T}\} = \mathcal{E}\{\mathbf{x}\mathbf{x}^{*T}\}\mathbf{D}^{*T} = \mathbf{R}_x\mathbf{D}^{*T}. \tag{2.43}$$

For two separate signals, x_1 and x_2, that are decimated as

$$y_1[m] = x_1[L_1 m] \qquad y_2[m] = x_2[L_2 m],$$

the cross-correlation matrix is given by

$$\mathbf{R}_{y_1 y_2} = \mathcal{E}\{\mathbf{y}_1\mathbf{y}_2^{*T}\} = \mathbf{D}_{L_1,M_1}\mathbf{R}_{x_1 x_2}\mathbf{D}_{L_2,M_2}^{*T}. \tag{2.44}$$

where the derivation for (2.44) is similar to that for (2.42) above.

The covariance matrix is defined as

$$\begin{aligned}\mathbf{C}_y &= \mathcal{E}\{(\mathbf{y}-\mathbf{m}_y)(\mathbf{y}-\mathbf{m}_y)^{*T}\} = \mathcal{E}\{(\mathbf{D}\mathbf{x}-\mathbf{D}\mathbf{m}_x)(\mathbf{D}\mathbf{x}-\mathbf{D}\mathbf{m}_x)^{*T}\} \\ &= \mathbf{D}\mathcal{E}\{(\mathbf{x}-\mathbf{m}_x)(\mathbf{x}-\mathbf{m}_x)^{*T}\}\mathbf{D}^{*T}\end{aligned}$$

or

$$\mathbf{C}_y = \mathbf{D}\mathbf{C}_x\mathbf{D}^{*T}. \tag{2.45}$$

The following relation can also be derived easily,

$$\mathbf{C}_y = \mathbf{R}_y - \mathbf{m}_y\mathbf{m}_y^{*T} = \mathbf{D}\left(\mathbf{R}_y - \mathbf{m}_y\mathbf{m}_y^{*T}\right)\mathbf{D}^{*T}.$$

2. Expansion

Now consider the expander defined by the input-output relation

$$y[m] = \begin{cases} x[m/L] & m \text{ div } L \\ 0 & \text{otherwise} \end{cases}.$$

The matrix form of expansion can be represented as

$$\mathbf{y} = \mathbf{U}_{L,M}\mathbf{x},$$

where the expansion matrix $\mathbf{U}$ is an $LM \times M$ matrix of 1's and 0's used to insert $L-1$ zeros between samples of $\mathbf{x}$ to form the output vector $\mathbf{y}$. This matrix is called an *expansion matrix*. For example, if the expansion factor is $L = 3$ and the length of the input vector is $M = 4$, then the associated expansion matrix is

$$\mathbf{U}_{3,4} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{2.46}$$

The expansion matrix can be represented as

$$\mathbf{U}_{L,M} = \mathbf{I}_M \otimes \mathbf{d}_L,$$

where $\mathbf{I}_M$ is the $M \times M$ identity matrix and $\mathbf{d}_l$ is a column vector defined in (2.40).

Using this definition and following the procedure used for decimation in the previous section, the matrix form of the output autocorrelation for the expander is

$$\mathbf{R}_y = \mathcal{E}\{\mathbf{y}\mathbf{y}^{*T}\} = \mathbf{U}\mathbf{R}_x\mathbf{U}^{*T}.$$

Likewise, the matrix form of the cross-correlation is

$$\mathbf{R}_{xy} = \mathcal{E}\{\mathbf{x}\mathbf{y}^{*T}\} = \mathbf{R}_x\mathbf{U}^{*T}.$$

The matrix results are of little use by themselves because there are large blocks of zeros as a result of the expansion operation. When expansion is followed by linear filtering however, the zeros fill in so a more meaningful correlation matrix results.

The cross-correlation of two interpolated signals,

$$y_1[m] = \begin{cases} x_1[m/L_1] & m \text{ div } L_1 \\ 0 & \text{otherwise} \end{cases}$$

$$y_2[m] = \begin{cases} x_2[m/L_2] & m \text{ div } L_2 \\ 0 & \text{otherwise} \end{cases}$$

is also of interest. The cross-correlation matrix for these two signals is

$$\mathbf{R}_{y_1y_2} = \mathcal{E}\{\mathbf{y}_1\mathbf{y}_2^{*T}\} = \mathbf{U}_{L_1,M_1}\mathbf{R}_{x_1x_2}\mathbf{U}_{L_2,M_2}^{*T}.$$

The covariance matrix for expansion follows a similar set of transformations. The covariance matrix of the expander is

$$\mathbf{C}_y = \mathbf{U}\mathbf{C}_x\mathbf{U}^{*T},$$

and satisfies the relation

$$\mathbf{C}_y = \mathbf{R}_y - \mathbf{m}_y\mathbf{m}_y^{*T} = \mathbf{U}\left(\mathbf{R}_x - \mathbf{m}_x\mathbf{m}_x^{*T}\right)\mathbf{U}^{*T}.$$

A direct relationship exists between the decimation matrix and the expansion matrix. The expansion matrix is the transpose of the decimation matrix,

$$\mathbf{U}_{L,M} = \mathbf{D}_{L,M}^T.$$

The proof of this result follows directly from the definition

$$\mathbf{U}_{L,M} = \mathbf{I}_M \otimes \mathbf{d}_L = \left(\mathbf{I}_M \otimes \mathbf{d}_L^T\right)^T = \mathbf{D}_{L,M}^T,$$

where property (3) of the Kronecker products was used (as described in Appendix B). This can also be seen in the examples of the decimation and expansion matrices, (2.39) and (2.46).

Some combinations of decimation and expansion are also worth mentioning. Expansion followed by decimation (using the same factor) is represented by the product matrix $\mathbf{D}_{L,M}\mathbf{U}_{L,M} = \mathbf{D}_{L,M}\mathbf{D}_{L,M}^T$. Since this pair of operations results in no change of the signal, it must follow that

$$\mathbf{D}_{L,M}\mathbf{U}_{L,M} = \mathbf{D}_{L,M}\mathbf{D}_{L,M}^T = \mathbf{I}_M. \tag{2.47}$$

On the other hand, decimation followed by expansion selects every L^{th} term of a sequence and replaces all other terms with zeros (the length of the sequence is unchanged). This operation is represented by the product

$$\mathbf{U}_{L,M}\mathbf{D}_{L,M} = \mathbf{D}_{L,M}^T\mathbf{D}_{L,M} = \mathbf{I}_M \otimes \mathbf{d}_L\mathbf{d}_L^T. \tag{2.48}$$

This product matrix is an orthogonal projection matrix. The geometric interpretation of this operation is a projection of the vector onto a subspace defined by every L^{th} component of the original vector.

3. Filters

The matrix form of a linear filter with finite support

$$y[m] = \sum_{r=P_-}^{P_+} h[r]x[m-r] \tag{2.49}$$

can be represented as

$$\mathbf{y} = \tilde{\mathbf{H}}^T\mathbf{x},$$

where, the term, $\tilde{\mathbf{H}}$, represents the reversal of the matrix $\mathbf{H}$, which is found by inverting the order of the elements along both the columns and the rows (see Appendix B.)[1]

For the case of a linear time-invariant filter, the matrix $\mathbf{H}$ has the form

$$\mathbf{H} = \begin{bmatrix} h[P_-] & 0 & \cdots & 0 & 0 \\ h[1+P_-] & h[P_-] & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ h[P_+] & h[P_+-1] & \cdots & 0 & 0 \\ 0 & h[P_+] & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & h[P_-] & 0 \\ 0 & 0 & \cdots & h[1+P_-] & h[P_-] \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & h[P_+] & h[P_+-1] \\ 0 & 0 & \cdots & 0 & h[P_+] \end{bmatrix}$$

[1]The reversal of a $P \times Q$ matrix $\mathbf{A}$ with elements a_{ij} is defined as a matrix $\tilde{\mathbf{A}}$ with elements $a_{P-i,Q-j}$.

where P_- and P_+ are the values of the lower and upper limits, respectively, of the filter in (2.49). Using this definition, the matrix form of the output autocorrelation of the filter is

$$\mathbf{R}_y = \mathcal{E}\{\mathbf{y}\mathbf{y}^{*T}\} = \tilde{\mathbf{H}}^T \mathbf{R}_x \tilde{\mathbf{H}}^*.$$

while, the matrix form of the cross-correlation is

$$\mathbf{R}_{xy} = \mathcal{E}\{\mathbf{x}\mathbf{y}^{*T}\} = \mathbf{R}_x \tilde{\mathbf{H}}^*.$$

If the lower limit of the filter matrix P_- equals zero, then the filter is causal, since the filter relies only on past observations of the input data. On the other hand, if $-P_-$ is less than zero, the filter uses future observations of the input. Therefore, the filter is non-causal and cannot be implemented in real-time. For an LPTV filter there are up to K possible sets of filter coefficients, $\mathbf{H}^{(k)}$. The appropriate set of filter coefficients is chosen based upon the system phase, as discussed earlier in this chapter.

F. SUMMARY

In this chapter, the fundamental building blocks and necessary indexing schemes for a multirate system were described. In addition, methods for calculating the correlations and power spectral densities for these building blocks were presented. Using these building blocks and relations, it is possible to develop the multirate Wiener-Hopf equations for optimal filtering of a multirate system. This is the topic of the next chapter.

THIS PAGE INTENTIONALLY LEFT BLANK

III. LINEAR ESTIMATION

In this section, the methodology for optimal filtering with multirate observations is developed. The basic principles of linear mean-square estimation apply to multirate signal processing; however, due to the periodic nature of the multirate system, one optimal filtering equation is not sufficient to fully describe the optimal filtering problem. By exploiting the system periodicity of the multirate system, a set of optimal filtering equations can be derived, one for each "phase" of the filter. The filter output (Wiener-Hopf equations) can be a single sequence or a vector of sequences based upon the input sequences.

A. SYSTEM EQUATIONS

To define the problem, assume a multirate observation model with a set of M wide-sense stationary observation sequences, $x_1[m_1] \ \ldots \ x_M[m_M]$. These sequences represent M observations of the signal $\overline{s}[n]$ subjected to various linear degradations (e.g., additive white Gaussian noise (WGN) or linear distortion), shown in Fig. 3.1. These observations sequences are to be used to estimate a single *desired* sequence $d[n]$, as shown in Fig. 3.2. Let $F_1, F_2, \ldots, F_M$ represent the sampling rates of the observation sequences and let $K_1, K_2, \ldots, K_M$ be the associated decimation factors. It will be assumed that the estimate $\hat{d}[n]$ is required at the fundamental rate $\overline{F}$, so the index 'n' is used for this sequence. The desired sequence at time n can be written as

$$d[n] = d[K \cdot l + k]$$

where K is the system period, $l = \lfloor n/K \rfloor$ and $n \equiv k \pmod{K}$. The variable k, is the sampling phase, as defined in Chapter II, with $0 \leq k \leq K-1$.

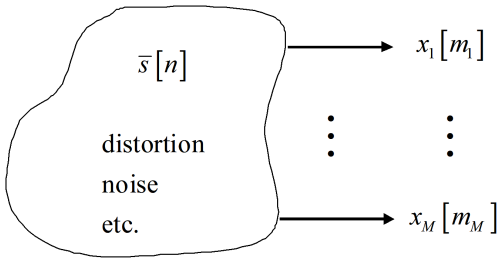


Figure 3.1. M-Channel Multirate Observation Model

Assume that a *causal* estimate is desired, i.e., the estimate is calculated using only the "present" and "past" observations. The optimal estimate is formed by summing the output of linear periodically time-varying (LPTV) filters, one for each observed sequence (see Fig. 3.2). Define the vector of filter coefficients

$$\mathbf{h}_i^{(k)} = \begin{bmatrix} h_i^{(k)}[0] \\ h_i^{(k)}[1] \\ \vdots \\ h_i^{(k)}[P_i - 1] \end{bmatrix} \qquad \begin{matrix} 1 \le i \le M \\ \\ 0 \le k \le K-1 \end{matrix}$$

where P_i is the order of the i^{th} filter and denote the vector of observations from the i^{th} channel by

$$\tilde{\mathbf{x}}_i^{(k)}[m_i] = \begin{bmatrix} x_i[m_i] & x_i[m_i - 1] & \cdots & x_i[m_i - P_i + 1] \end{bmatrix}^T,$$

where $m_i = \lfloor n/K_i \rfloor$. Here m_i represents the most recent observation time for x_i if the causality constraint is to be maintained. Then the estimate at time 'n' is given by

$$\hat{d}_k[n] = \sum_{i=1}^{M} \tilde{\mathbf{x}}_i^{(k)T}[n]\, \mathbf{h}_i^{(k)} = \sum_{i=1}^{M} \mathbf{h}_i^{(k)T} \tilde{\mathbf{x}}_i^{(k)}[n]; \quad \text{where } n \equiv k \pmod{K}. \tag{3.1}$$

The subscript k in $\hat{d}_k[n]$ is actually redundant since $n \equiv k \pmod{K}$; however, this notation is used to emphasize that the statistical properties of the estimate are periodically time varying.

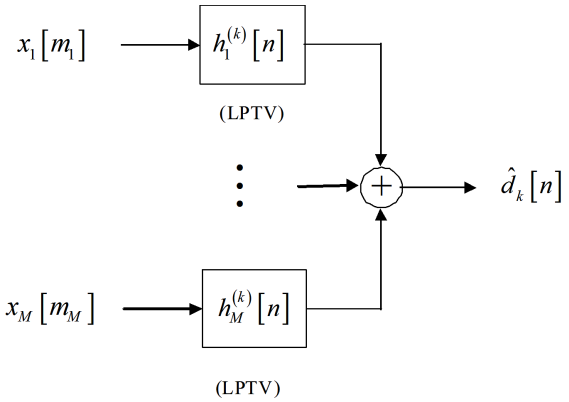


Figure 3.2. Multirate Optimal Filter (Direct Form)

The form of the estimation given by (3.1) and depicted in Fig. 3.2 will be called the "direct form" realization since the input sequences are directly weighed and summed. In the next section, the methods for finding the optimal filter parameters in the direct form and the corresponding mean-square error variance are developed.

B. MULTIRATE WIENER-HOPF EQUATIONS

Having defined the estimate, it is now possible to define the error as

$$\varepsilon_k[n] = d[n] - \hat{d}_k[n] \tag{3.2}$$

and to find the optimal set of filter coefficients that minimize $\mathcal{E}\{|\varepsilon_k[n]|^2\}$. The orthogonality principle of linear mean-square estimation [Ref. 18, 58] requires that the error be orthogonal to the observation vectors, i.e.,

$$\mathcal{E}\{\mathbf{x}_i^{(k)}[n]\varepsilon_k^*[n]\} = 0; \qquad i = 1, 2, \ldots, M. \tag{3.3}$$

Substituting (3.1) and (3.2) into (3.3) and taking the expectation yields

$$\mathcal{E}\{\tilde{\mathbf{x}}_i^{(k)}[n]\varepsilon_k^*[n]\} = E\left\{\tilde{\mathbf{x}}_i^{(k)}[n]\left(d[n] - \sum_{j=1}^{M}\tilde{\mathbf{x}}_j^{(k)T}[n]\,\mathbf{h}_j^{(k)}\right)^*\right\} = \tilde{\mathbf{r}}_{di}^{(k)*} - \sum_{j=1}^{M}\tilde{\mathbf{R}}_{ij}^{(k)}[n]\,\mathbf{h}_j^{(k)*} = 0$$

where the following terms have been defined:

$$\tilde{\mathbf{r}}_{di}^{(k)*}[n] = \mathcal{E}\{\tilde{\mathbf{x}}_i^{(k)}[n]d^*[n]\} \tag{3.4}$$

$$\tilde{\mathbf{R}}_{ij}^{(k)}[n] = \mathcal{E}\{\tilde{\mathbf{x}}_i^{(k)}[n]\tilde{\mathbf{x}}_j^{(k)*T}[n]\}. \tag{3.5}$$

This can be rearranged and written as

$$\tilde{\mathbf{r}}_{di}^{(k)*} = \sum_{j=1}^{M} \tilde{\mathbf{R}}_{ij}^{(k)}[n]\, \mathbf{h}_j^{(k)*}, \quad i = 1, 2, \ldots, M$$

or equivalently as

$$\tilde{\mathbf{r}}_{di}^{(k)} = \sum_{j=1}^{M} \tilde{\mathbf{R}}_{ij}^{(k)*}[n]\, \mathbf{h}_j^{(k)}, \quad i = 1, 2, \ldots, M. \tag{3.6}$$

Then using the Hermitian symmetry property $\tilde{\mathbf{R}}_{ij}^{(k)*} = \tilde{\mathbf{R}}_{ji}^{(k)T}$, (3.6) can be represented in matrix form as

$$\begin{bmatrix} \tilde{\mathbf{R}}_{11}^{(k)} & \tilde{\mathbf{R}}_{12}^{(k)*} & \cdots & \tilde{\mathbf{R}}_{1M}^{(k)*} \\ \tilde{\mathbf{R}}_{12}^{(k)T} & \tilde{\mathbf{R}}_{22}^{(k)} & \cdots & \tilde{\mathbf{R}}_{2M}^{(k)*} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\mathbf{R}}_{1M}^{(k)T} & \tilde{\mathbf{R}}_{2M}^{(k)T} & \cdots & \tilde{\mathbf{R}}_{MM}^{(k)} \end{bmatrix} \begin{bmatrix} \mathbf{h}_1^{(k)} \\ \mathbf{h}_2^{(k)} \\ \vdots \\ \mathbf{h}_M^{(k)} \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{r}}_{d1}^{(k)} \\ \tilde{\mathbf{r}}_{d2}^{(k)} \\ \vdots \\ \tilde{\mathbf{r}}_{dM}^{(k)} \end{bmatrix}, \tag{3.7}$$

The associated error variance according to the orthogonality principle is given by $\sigma_k^2 = \mathcal{E}\{d[n]\varepsilon_k^*[n]\}$ which, upon substitution of (3.1) and (3.2) becomes

$$\sigma_k^2 = R_d(0) - \sum_{i=1}^{M} \tilde{\mathbf{r}}_{di}^{(k)*T} \mathbf{h}_i^{(k)}. \tag{3.8}$$

Observe that the mean-square error σ_k^2 is periodically time varying, i.e., it is dependent on the phase k of the filter. In order to establish a single figure of merit for the system, an average error is defined,

$$\sigma_\varepsilon^2 = \left(\prod_{k=0}^{K-1} \sigma_k^2 \right)^{\frac{1}{K}} \tag{3.9}$$

This geometrical average is chosen, since the error, when expressed in decibels (dB) becomes the arithmetic average of the σ_k^2 values in dB. For Gaussian signals, this relates to average information (entropy).

C. CALCULATION OF CORRELATION TERMS

The correlation terms needed in (3.7) and (3.8) can most easily be generated by using the linear algebra concepts described in Chapter II. For any time n and corresponding index k, the observation sequence $\mathbf{x}_i^{(k)}[n]$ can be expressed as

$$\tilde{\mathbf{x}}_i^{(k)}[n] = \tilde{\mathbf{D}}_{L_i}^{(k)} \tilde{\mathbf{x}}_i[n], \tag{3.10}$$

where $\tilde{\mathbf{x}}_i$ is given by

$$\tilde{\mathbf{x}}_i[n] = \begin{bmatrix} \overline{x}_i[n] & \overline{x}_i[n-1] & \cdots & \overline{x}_i[n - P_i \cdot K_i + 1] \end{bmatrix}^T$$

and $\tilde{\mathbf{D}}_{L_i}^{(k)}$ is an appropriately-defined decimation matrix. Note that $\tilde{\mathbf{x}}_i[n]$ consists of all possible values of $x_i[n]$ (observed or unobserved) and $\tilde{\mathbf{x}}_i^{(k)}[n]$ represents just the observed values used by the filter.

By virtue of (3.10), the correlation matrix $\tilde{\mathbf{R}}_{ij}^{(k)}$ of $\tilde{\mathbf{x}}_i[n]$ and $\tilde{\mathbf{x}}_j[n]$ is given by

$$\tilde{\mathbf{R}}_{ij}^{(k)} = \tilde{\mathbf{D}}_{L_i}^{(k)} \tilde{\mathbf{R}}_{ij} \tilde{\mathbf{D}}_{L_j}^{(k)T}, \tag{3.11}$$

where $\tilde{\mathbf{R}}_{ij}$ is the correlation matrix of $\tilde{\mathbf{x}}_i[n]$ and $\tilde{\mathbf{x}}_j[n]$ at the fundamental layer

$$\tilde{\mathbf{R}}_{ij} = \mathcal{E}\{\tilde{\mathbf{x}}_i[n] \tilde{\mathbf{x}}_j^{*T}[n]\}.$$

The cross-correlation vector between $\tilde{\mathbf{x}}_i$ and $\mathbf{d}$ is

$$\tilde{\mathbf{r}}_{di}^{(k)} = \tilde{\mathbf{D}}_{L_i}^{(k)} \tilde{\mathbf{r}}_{di}. \tag{3.12}$$

where $\tilde{\mathbf{r}}_{di}$ is the cross-correlation of between $\tilde{\mathbf{x}}_i$ and $\mathbf{d}$ at the fundamental layer

$$\tilde{\mathbf{r}}_{di} = \mathcal{E}\{\tilde{\mathbf{x}}_i[n]d^*[n]\}.$$

Note that the dependency on the system phase is completely determined by the decimation matrices.

D. FILTER COEFFICIENTS

The solution to the multirate Wiener-Hopf equations produces the parameters for the *direct form* realization of the filter. An alternative realization can be developed that leads to some useful insights about the role that each input plays in forming the overall estimate $\hat{d}_k[n]$. This alternative realization is a recursive form of the optimal filter equations, and a complete derivation of these equations is contained in Appendix C. It is shown there that the filter coefficients for the direct form can be written as:

$$\mathbf{h}_i^{(k)} = \mathbf{E}_i^{(k)^{-1}} \left(\tilde{\mathbf{r}}_{di}^{(k)} - \mathbf{G}_i^{(k)*T} \tilde{\mathbf{b}}_{di}^{(k)} \right) - \sum_{j=i+1}^{M} \mathbf{E}_i^{(k)^{-1}} \left(\tilde{\mathbf{R}}_{ij}^{(k)} - \mathbf{G}_i^{(k)*T} \tilde{\mathbf{B}}_{ij}^{(k)} \right) \mathbf{h}_j^{(k)} \quad (1 \leq i \leq M-1) \tag{3.13}$$

$$\mathbf{h}_M^{(k)} = \mathbf{E}_M^{(k)^{-1}} \left(\tilde{\mathbf{r}}_{dM}^{(k)} - \mathbf{G}_M^{(k)*T} \tilde{\mathbf{b}}_{dM}^{(k)} \right) \tag{3.14}$$

where the vector $\tilde{\mathbf{b}}_{di}^{(k)}$ and the matrix $\tilde{\mathbf{B}}_{ij}^{(k)}$ are defined as

$$\tilde{\mathbf{b}}_{di}^{(k)} = \begin{bmatrix} \tilde{\mathbf{r}}_{d1}^{(k)} \\ \vdots \\ \tilde{\mathbf{r}}_{d(i-1)}^{(k)} \end{bmatrix} \qquad \tilde{\mathbf{B}}_{ij}^{(k)} = \begin{bmatrix} \tilde{\mathbf{R}}_{1j}^{(k)} \\ \vdots \\ \tilde{\mathbf{R}}_{(i-1)j}^{(k)} \end{bmatrix}$$

and the matrices $\mathbf{G}_i^{(k)}$ and $\mathbf{E}_i^{(k)}$ are given by

$$\mathbf{G}_i^{(k)} = \begin{cases} 0 & i = 1 \\ \begin{bmatrix} \tilde{\mathbf{R}}_1^{(k)} & \tilde{\mathbf{R}}_{12}^{(k)} & \cdots & \tilde{\mathbf{R}}_{1(i-1)}^{(k)} \\ \tilde{\mathbf{R}}_{12}^{(k)*T} & \tilde{\mathbf{R}}_2^{(k)} & \cdots & \tilde{\mathbf{R}}_{2(i-1)}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\mathbf{R}}_{1(i-1)}^{(k)*T} & \tilde{\mathbf{R}}_{2(i-1)}^{(k)*T} & \cdots & \tilde{\mathbf{R}}_{(i-1)}^{(k)} \end{bmatrix}^{-1} \tilde{\mathbf{B}}_{ii}^{(k)} & 1 < i \leq M \end{cases}$$

$$\mathbf{E}_i^{(k)} = \begin{cases} \tilde{\mathbf{R}}_{11}^{(k)} & i = 1 \\ \tilde{\mathbf{R}}_{ii}^{(k)} - \tilde{\mathbf{G}}_i^{(k)*T} \mathbf{B}_{ii}^{(k)} & 1 < i \leq M \end{cases}$$

The filter coefficient, $\mathbf{h}_i^{(k)}$ in (3.13), can then be re-expressed as

$$\mathbf{h}_i^{(k)} = \mathbf{h}_i'^{(k)} - \sum_{j=i+1}^{M} \mathbf{H}_{ij}^{(k)} \mathbf{h}_j^{(k)}, \tag{3.15}$$

where

$$\mathbf{h}_i'^{(k)} = \mathbf{E}_i^{(k)^{-1}} \left(\tilde{\mathbf{r}}_{di}^{(k)} - \mathbf{G}_i^{(k)*T} \tilde{\mathbf{b}}_{di}^{(k)} \right)$$

and

$$\mathbf{H}_{ij}^{(k)} = \mathbf{E}_i^{(k)^{-1}} \left(\tilde{\mathbf{R}}_{ij}^{(k)} - \mathbf{G}_i^{(k)*T} \tilde{\mathbf{B}}_{ij}^{(k)} \right).$$

With these definitions it can be seen from (3.15) that the filter $\mathbf{h}_i^{(k)}$ is comprised of a modified optimal filter, $\mathbf{h}_i'^{(k)}$, and cross (or prediction) filters, $\mathbf{H}_{ij}^{(k)}$, which allows the optimal filter to be represented in the form shown in Fig. 3.3. This form is referred to as the *innovations representation* because each branch of the realization works on just the new information not present in the other existing channels. It can be seen that after applying the cross filters a modified set of observations $\{x_1' \cdots x_M'\}$ is obtained. The modified observation x_i' is the residual after predicting x_i using x_j, $1 \leq j \leq i-1$ and represents the new information brought in by channel i. The primed observations are mutually orthogonal. The modified optimal filtering terms,

$\mathbf{h}'_i$, $i = 1, 2, \ldots, j$, represent the optimal filter for estimating d if $x_j[m_j]$, $j > i$ is not used.

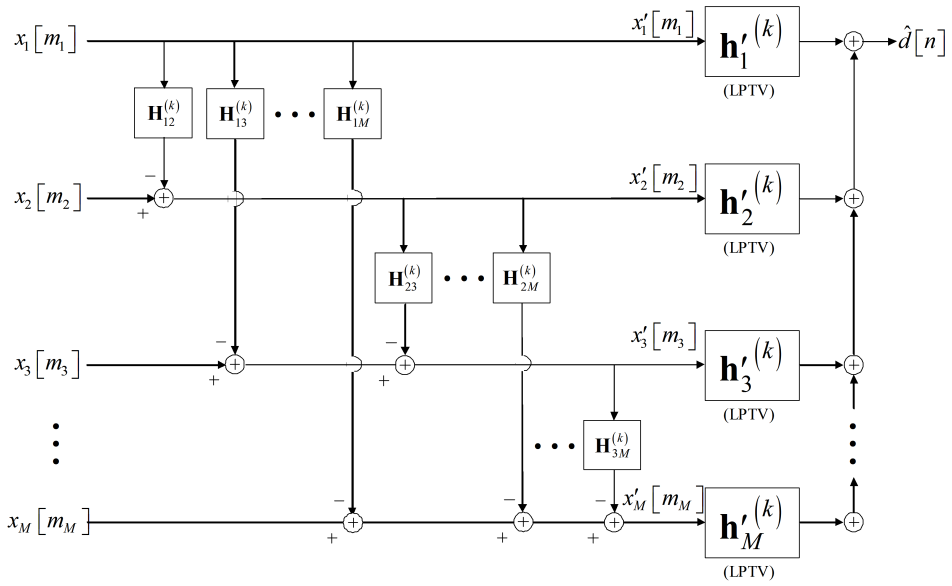


Figure 3.3. Multirate Innovations Representation

By separating the filter into cross filter terms and a modified optimal filtering term for each input signal, it is possible to determine the improvement in error variance that would be derived from adding extra signals, without having to recalculate all of the filter coefficients. For each new signal added only the prior filtering terms and the modified optimal filtering terms associated with the new signal are calculated. All the previous prior filtering and modified optimal filtering terms remain unaffected.

By substituting the filter coefficient definition of (3.15) into the error variance of (3.8) the error variance can be expressed in a recursive form dependent on the number of signals observed. Let $\sigma^2_{k,j}$ be the phase-periodic error variance when only j signals are observed. This form of the error variance will be called the innovations form of the error variance equation and is expressed as

$$\sigma_{k,M}^2 = \sigma_{k,M-1}^2 - \tilde{\mathbf{r}}_{dM}^{(k)\,T}\mathbf{h}_M^{(k)*} - \sum_{i=1}^{M-1}\tilde{\mathbf{r}}_{dM}^{(k)\,T}\left(-\mathbf{H}_{iM}^{(k)*}\right)\mathbf{h}_{i_M}^{'(k)*}$$
$$- \sum_{i=1}^{M-1}\sum_{i_1=i+1}^{M}\tilde{\mathbf{r}}_{dM}^{(k)\,T}\left(-\mathbf{H}_{ii_1}^{(k)*}\right)\left(-\mathbf{H}_{i_1M}^{(k)*}\right)\mathbf{h}_M^{'(k)*} - \cdots$$
$$- \sum_{i=1}^{M-1}\sum_{i_1=i+1}^{M}\sum_{i_2=i_1+1}^{M}\cdots\sum_{i_{M-1}=i_{M-2}+1}^{M}\tilde{\mathbf{r}}_{dM}^{(k)\,T}\left(-\mathbf{H}_{ii_1}^{(k)*}\right)\times$$
$$\left(-\mathbf{H}_{i_1i_2}^{(k)*}\right)\cdots\left(-\mathbf{H}_{i_{M-2}i_{M-1}}^{(k)*}\right)\left(-\mathbf{H}_{i_{M-1}M}^{(k)*}\right)\mathbf{h}_{i_M}^{'(k)*}. \quad (3.16)$$

The derivation for this form is also contained in Appendix C. Though this form of the error variance can become unwieldly for systems with a large number of observed signals, this equation is useful to show the improvement in performance brought about by each additional channel.

E. PERFORMANCE STUDY

A typical scenario for multirate estimation is based on the signal model shown in Fig. 3.4. The signal of interest is $\overline{s}[n]$, therefore $d[n] = \overline{s}[n]$. The desired signal is subject to independent additive noise in each channel before decimation. In particular, this study considers a system with two observation sequences $x_1[m_1]$ and $x_2[m_2]$, with decimation factors of $K_1 = 2$ and $K_2 = 5$.

Two different types of signals were analyzed for comparison: a second order AR process and a periodic signal comprised of two sinusoids. Both signals are defined in the fundamental layer by the equations shown in Table 3.1; where the driving term $\overline{w}[n]$ in the AR process is white Gaussian noise (WGN) with mean zero and unit variance.

The associated autocorrelation functions for each of the signals were calculated and are provided in Table 3.2.

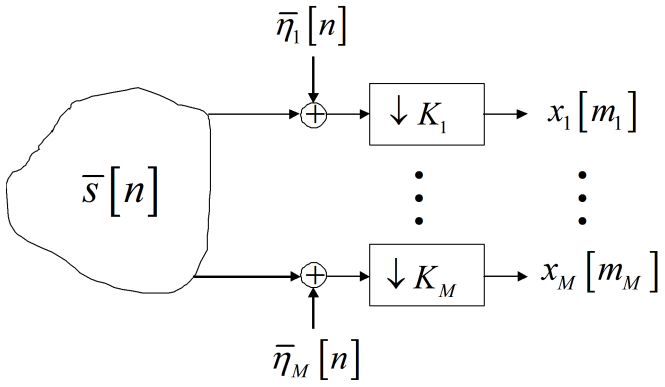


Figure 3.4. Multirate model with multiple observation sequences

Table 3.1. Signal Examples

Signal Type	Signal Equation
2^{nd} Order AR	$\overline{s}[n] = 0.9\overline{s}[n-1] - 0.8\overline{s}[n-2] + \overline{w}[n]$
2 Sinusoids	$\overline{s}[n] = 4\cos[0.2\pi n] + 2\cos[0.02\pi n]$

The variances of the two noise sequences η_1 and η_2 were chosen such that the Signal to Noise Ratio (SNR) assumed the values, -6 dB, -3 dB, -1.7 dB, 0 dB, 1.7 dB, 3 dB and 6 dB. The corresponding noise variance can be calculated from

$$\sigma_\eta^2 = \frac{\mathrm{R}_s(0)}{10^{(\mathrm{SNR}/10)}}$$

where R_s is the autocorrelation function for $\overline{s}[n]$. A block diagram of the optimal estimator along with the sampling pattern for the observations is shown in Fig. 3.5. Initial simulations were conducted with filter orders of $P_1 = P_2 = 5$.

Figure 3.6 shows the theoretical error variances for the second order AR model, computed using (3.8) and (3.9). The figure compares the error variances for using the low-rate signal alone, for using the high-rate signal alone, and for using the low- and high-rate signals in combination. In this particular experiment, the high-rate signal, x_1, was observed in a 0 dB noise environment and the low-rate signal, x_2 was observed with a SNR varying over the range -6 to 6 dB. From this figure, one can

Table 3.2. Signal Autocorrelation Functions

Signal Type	Autocorrelation Function
2^{nd} order AR	$\mathrm{R}_s(l) = \begin{cases} 3.7114 \cdot (0.8944)^l \cos(1.0436l - 0.0646), & l \geq 0 \\ -3.7114 \cdot (1.1180)^l \cos(1.0436l - 3.0770), & l < 0 \end{cases}$
2 sinusoids	$\mathrm{R}_s(l) = 8\cos(0.2\pi l) + 2\cos(0.02\pi l),\ \forall l$

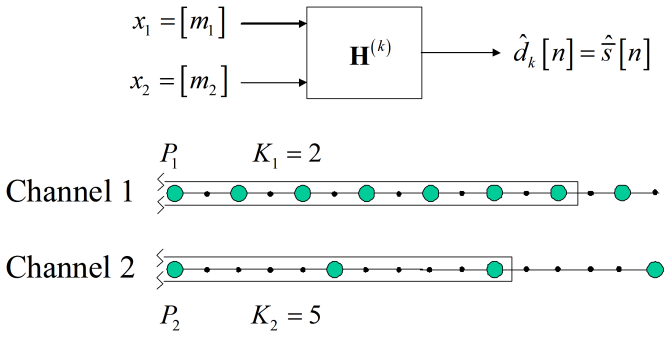


Figure 3.5. Optimal Estimator Block Diagram

see that the error variance for the low-rate signal, over the SNR range of -6 dB to $+6$ dB, is higher than the error variance of the high rate signal observed at 0 dB. Individually, the high-rate signal performs better than the low rate signal. When the low-rate and high-rate signals are processed together, however, the error variance is lower than for either signal individually. At the lowest SNR simulated (-6 dB) only a modest improvement was observed. As the SNR of the low rate signal increases, however, the improvement in the error variance approaches 2.5 dB.

Figure 3.7 shows the theoretical error variances for the sinusoidal model. Again, the high-rate signal, x_1, was observed in a 0 dB noise environment and the low-rate signal, x_2 was observed with a SNR varying over the range -6 dB to $+6$ dB. Again in this case, the error variance for the low-rate signal at all SNRs tested is higher than the error variance of the high-rate signal at 0 dB. When the low-rate and high-rate signals are processed together, however, the error variance is lower than

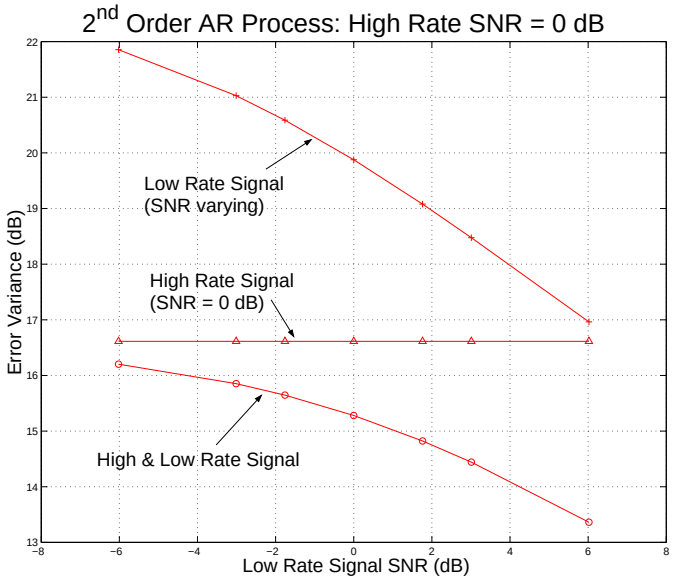


Figure 3.6. Error Variance vs SNR for the 2nd Order AR Process

for either signal individually. At the lowest SNR simulated (−6 dB) only a modest improvement was observed. As the SNR of the low-rate signal increases, however, the improvement in the error variance approaches 6 dB.

Additional simulations were conducted with filter orders P_1 and P_2 ranging from 5 to 30 for an environment where the SNR for both the low rate and high rate signals is 6 dB. Selected results are listed in Table 3.3 for the signal with two sinusoids, showing the performance that results when using both the high rate and the low rate observations together or using only one set of observations at a time. From Table 3.3, one can see that the use of both sets of observations results in significant improvement (3 dB to 5 dB) over using either x_1 or x_2 separately, although the error associated with filtering using only x_2 is large compared to that resulting from using only x_1.

The effect on order for the AR process could not be studied. Because of the low order of the AR process, varying the filter orders between 5 and 30 has no appreciable

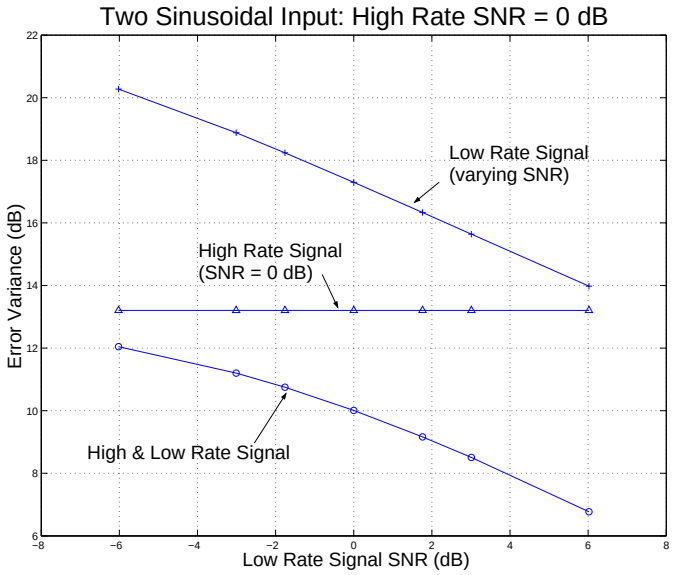


Figure 3.7. Error Variance vs SNR for the Periodic Signal

effect on the error variance.

F. SUMMARY

This chapter addressed the problem of optimal filtering of multiple related channels of data observed at different sampling rates. The direct form of the Wiener-Hopf equations using linear, periodically time-varying filters was developed, and the error variance of the optimal filter was observed to be periodically time-varying. In addition to the direct form, an innovations form of the multirate Wiener-Hopf equations was presented. This recursive form provides insight into the relative change in performance when additional signals are added or removed from the optimal filter. Using the geometric average of the periodic error variances as a performance measure, it was demonstrated that optimal filtering of multiple channels can provide improved performance over optimal filtering using one channel, even if the secondary channel when used alone has high error variances. In the simulations conducted, the

Table 3.3. Error Variances for Periodic Signals with 2 sinusoids (All values in dB)

Filter Orders	σ_ε^2 $x_1[n]$ and $x_2[n]$	σ_ε^2 $x_1[n]$ only	σ_ε^2 $x_2[n]$ only
$P = 5, Q = 5$	-0.43	3.27	13.98
$P = 30, Q = 5$	-13.54	-12.26	13.98
$P = 5, Q = 30$	-6.54	3.27	9.39
$P = 30, Q = 30$	-17.32	-12.26	9.39

largest improvements occured when the signal to be estimated consisted of sinusoids in noise. When the signal was derived from a second order AR process, the use of multiple channels and/or higher order filters resulted in only a small improvement in performance. It was observed that a single channel filter of order 5 was sufficient to estimate the second order AR signal in noise. These examples used only two observation sequences, but the derivations and experiments presented here can be applied to an arbitrary number of channels.

IV. LINEAR PREDICTION

The previous chapter discussed estimation of data from noisy signals using present observations and a finite number of past observations. Applications that require the ability to estimate the present value of the observations themselves using only past observations (causal filtering), however, exist. A few of the most common applications are focused on speech and image coding, target tracking and target classification. The process of estimating a signal using only its past observations is known as *linear prediction*. Linear prediction is at the core of all linear estimation problems, such as the more general Wiener filter. Furthermore, specific insights that arise from linear prediction, such as the lattice forms of the filter, can be carried over to the more general problems.

A. SINGLE-CHANNEL AND MULTICHANNEL REVIEW

Linear prediction has been well researched for the single-channel and multichannel cases. At its most intrinsic form, the goal is to estimate the present value of a data sequence from a finite collection of past data. For the single-channel case, the prediction equation can be written

$$\begin{aligned}\hat{x}[n] &= -a_1^* x[n-1] - a_2^* x[n-2] - \cdots - a_P^* x[n-P] \\ &= -\sum_{i=1}^{P} a_i^* x[n-i]\end{aligned}$$

where $\hat{x}[n]$ is the predicted value of $x[n]$ using $x[n-1]$ through $x[n-P]$ and a_1 through a_P are the prediction filter coefficients (written as conjugate negative values for later convenience). The value P is referred to as the prediction filter order.

For the multichannel case, the prediction equation generalizes to

$$\begin{aligned}\hat{\mathbf{x}}[n] &= -\mathbf{A}_1^{*T}\mathbf{x}[n-1] - \mathbf{A}_2^{*T}\mathbf{x}[n-2] - \cdots - \mathbf{A}_P^{*T}\mathbf{x}[n-P] \\ &= -\sum_{i=1}^{P} \mathbf{A}_i^{*T}\mathbf{x}[n-i] \end{aligned} \tag{4.1}$$

where the vector $\mathbf{x}$ represents the data observed in each of M channels and $\mathbf{A}_1$ through $\mathbf{A}_P$ are the prediction filter coefficient matrices. This more general case will now be discussed.

The goal of the linear prediction problem is to find an optimal solution for the coefficients $\mathbf{A}_i$ in (4.1), such that the expected value of the squared norm of the prediction error $\boldsymbol{\varepsilon}[n] = \mathbf{x}[n] - \hat{\mathbf{x}}[n]$ is minimized. The solution to the linear prediction problem results in the well known (multichannel) Normal equations

$$\begin{bmatrix} \mathbf{R}[0] & \mathbf{R}[1] & \cdots & \mathbf{R}[P] \\ \mathbf{R}[-1] & \mathbf{R}[0] & \cdots & \mathbf{R}[P-1] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}[-P] & \mathbf{R}[-P+1] & \cdots & \mathbf{R}[0] \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_P \end{bmatrix} = \begin{bmatrix} \mathbf{E} \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \tag{4.2}$$

where the multichannel correlation function is defined as $\mathbf{R}[i] = \mathcal{E}\{\mathbf{x}[n]\mathbf{x}^{*T}[n-i]\}$ and $\mathbf{E}$ is the prediction error covariance matrix $\mathbf{E} = \mathcal{E}\{\boldsymbol{\varepsilon}[n]\boldsymbol{\varepsilon}^{*T}[n]\}$.

If the order of the prediction filter is allowed to grow at each successive observation, in order to include all past observations at each step, then the calculated prediction errors $\boldsymbol{\varepsilon}[n]$ will be orthogonal. If a prediction filter of sufficiently high order is used, then the calculated prediction errors will be approximately orthogonal ($\mathcal{E}\{\boldsymbol{\varepsilon}[i]\boldsymbol{\varepsilon}^{*T}[j] = 0\}$ for $i \neq j$). Thus the prediction error filter can be thought of as

a causal whitening filter. This property is important in autoregressive (AR) modeling, moving average (MA) modeling and autoregressive moving average (ARMA) modeling (see [Ref. 18, 58, 19].)

B. MULTIRATE LINEAR PREDICTION THEORY

The method for extending linear prediction to a multirate system will now be considered. Recall from Chapter II, that multirate systems are periodic in nature and that one can determine an associated system period K. This system period can be used to partition the data into blocks with a common sampling structure as illustrated in Fig. 4.1

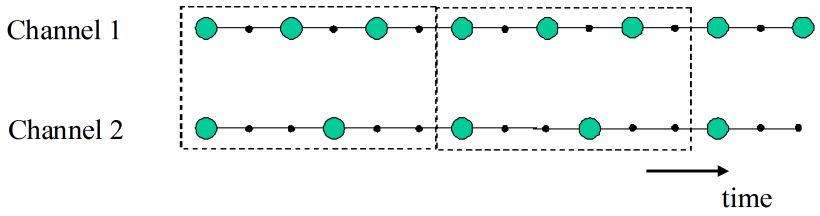


Figure 4.1. Multirate System Block Structure

In this example, Channel 1 is observed at half the fundamental rate, i.e., the decimation factor is 2, while for the Channel 2 the decimation factor is 3. The system period, as defined in Chapter II, for this example is $K = 6$. This has been highlighted in Fig. 4.1 by the dashed boxes representing blocks of data. When viewed at the fundamental rate, the statistics of the multirate system are cyclo-stationary. However, when the multirate system is viewed in blocks, the block statistics are stationary.

In the specific linear prediction problem that is considered here, each new data point (in any channel) is predicted as it occurs in time. The prediction is based on a finite number Q of previous full blocks of data as well as all of the data within the current block that occurs earlier than the point being predicted. Specifically, let m

represent a time index for the blocks and $i = 1, 2, \ldots, P$ be an index for the data occurring within a block, also ordered in time. For the moment, assume that data points from different channels do *not* occur at exactly the same time. The prediction equation for the $(i+1)^{st}$ data point is then written as

$$\begin{aligned}\hat{x}_i[m] &= -\alpha^*_{i,1}x_{i-1}[m] - \cdots - \alpha^*_{i,i-1}x_1[m] - \mathbf{a}^{*T}_{i,1}\mathbf{x}[m-1] - \cdots - \mathbf{a}^{*T}_{i,Q}\mathbf{x}[m-Q] \\ &= -\sum_{p=1}^{i-1}\alpha^*_{i,p}x_{i-p}[m] - \sum_{q=1}^{Q}\mathbf{a}^{*T}_{i,q}\mathbf{x}[m-q]. \end{aligned} \tag{4.3}$$

The variable $x_i[m]$ represents the i^{th} observed data point for the m^{th} system block, and $\mathbf{x}[m]$ is a vector of all the observed data associated with the m^{th} system block. The elements of $\mathbf{x}[m]$ are

$$\mathbf{x}[m] = \begin{bmatrix} x_P[m] \\ x_{P-1}[m] \\ \vdots \\ x_1[m] \end{bmatrix}. \tag{4.4}$$

This method of ordering the block data points is for indexing purposes only; it does not account for data being observed at the same time. The data is ordered according to oldest data first with channel 1 being considered the oldest channel for indexing purposes. Figure 4.2 shows these variables for the system depicted in Fig. 4.1.

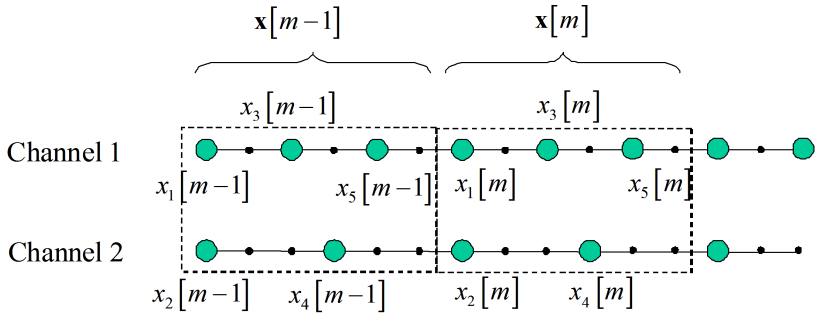


Figure 4.2. Multirate System Block Variables

The prediction error for $x_i[m]$ is given by

$$\varepsilon_i[m] = x_i[m] - \hat{x}_i[m]$$

or

$$\varepsilon_i[m] = x_i[m] + \sum_{p=1}^{i-1} \alpha_{i,p}^* x_{i-p}[m] + \sum_{q=1}^{Q} \mathbf{a}_{i,q}^{*T} \mathbf{x}[m-q] \tag{4.5}$$

By applying the principle of orthogonality [Ref. 18], one can determine the prediction error variance for each predicted value and the set of equations that the prediction coefficients must satisfy. The prediction error variance satisfies

$$\sigma_i^2 = \mathcal{E}\{x_i[m]\varepsilon_i^*[m]\} = R_{x_i}[0] + \sum_{p=1}^{i-1} \alpha_{i,p} R_{x_i x_{i-p}}[0] + \sum_{q=1}^{Q} \mathbf{a}_{i,q}^T \mathbf{r}_{x_i \mathbf{x}}[q] \tag{4.6}$$

where $\mathbf{r}_{x_i \mathbf{x}}[l]$ is the vector of cross-correlation terms

$$\mathbf{r}_{x_i \mathbf{x}}[l] = \begin{bmatrix} R_{x_i x_P}[l] \\ R_{x_i x_{P-1}}[l] \\ \vdots \\ R_{x_i x_1}[l] \end{bmatrix}.$$

Setting the observations orthogonal to the prediction error produces the equations

$$\begin{aligned} 0 &= \mathcal{E}\{x_j[m-l]\varepsilon_i^*[m]\} \\ &= R_{x_j x_i}[-l] + \sum_{p=1}^{P} \alpha_{i,p} R_{x_j x_{i-p}}[-l] + \sum_{q=1}^{Q} \mathbf{a}_{i,q}^T \mathbf{r}_{x_j \mathbf{x}}[q-l]. \end{aligned} \tag{4.7}$$

Combining these sets of equations for all observations within a data block, and using the shorthand notation

$$\mathbf{R}_l = \mathbf{R}_{\mathbf{x}}[l] = \mathcal{E}\left[\mathbf{x}[m]\mathbf{x}^{*T}[m-l]\right], \tag{4.8}$$

produces the matrix form of the *multirate* Normal equations

$$\begin{bmatrix} \mathbf{R}_0 & \mathbf{R}_1 & \cdots & \mathbf{R}_Q \\ \mathbf{R}_{-1} & \mathbf{R}_0 & \cdots & \mathbf{R}_{Q-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{-Q} & \mathbf{R}_{1-Q} & \cdots & \mathbf{R}_0 \end{bmatrix} \begin{bmatrix} \mathbf{A}_0 \\ \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_Q \end{bmatrix} = \begin{bmatrix} \mathbf{E} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \tag{4.9}$$

where

$$\mathbf{A}_0 = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \alpha_{P,1} & 1 & \cdots & 0 \\ \alpha_{P,2} & \alpha_{P-1,1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{P,P-1} & \alpha_{P-1,P-2} & \cdots & 1 \end{bmatrix} \tag{4.10}$$

$$\mathbf{A}_i = \begin{bmatrix} | & | & & | \\ \mathbf{a}_{P,i} & \mathbf{a}_{P-1,i} & \cdots & \mathbf{a}_{1,i} \\ | & | & & | \end{bmatrix} \qquad i = 1, 2, \ldots, Q, \tag{4.11}$$

and the error covariance matrix is

$$\mathbf{E} = \begin{bmatrix} \sigma_P^2 & \times & \cdots & \times \\ 0 & \sigma_{P-1}^2 & \cdots & \times \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_1^2 \end{bmatrix}. \tag{4.12}$$

The matrix elements represented by an '$\times$' do not affect the linear prediction problem, so their values do not have to be calculated.

So far, it has been assumed that data points from different channels do not occur at the same time. If points *do* occur at the same time then they must be jointly predicted. For instance, for the system depicted in Figure 4.2, data points occur in both channels at the first observation time of the data block. The prediction equation for $x_1[m]$ is of the form

$$\hat{x}_1[m] = -\sum_{q=1}^{Q} \mathbf{a}_{1,q}^{*T} \mathbf{x}[m-q].$$

This equation does not change. The linear prediction equation of (4.3) for $x_2[m]$, however, would change from

$$\hat{x}_2[m] = -\alpha^*_{2,1} x_1[m] - \sum_{q=1}^{Q} \mathbf{a}^{*T}_{2,q} \mathbf{x}[m-q]$$

to

$$\hat{x}_2[m] = -\sum_{q=1}^{Q} \mathbf{a}^{*T}_{2,q} \mathbf{x}[m-q].$$

In other words, $x_1[m]$ and $x_2[m]$ are predicted simultaneously. The prediction coefficient matrix and prediction error matrix of (4.10) and (4.12) for the 5×5 system with no joint observations are

$$\mathbf{A}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ \alpha_{5,1} & 1 & 0 & 0 & 0 \\ \alpha_{5,2} & \alpha_{4,1} & 1 & 0 & 0 \\ \alpha_{5,3} & \alpha_{4,2} & \alpha_{3,1} & 1 & 0 \\ \alpha_{5,4} & \alpha_{4,3} & \alpha_{3,2} & \alpha_{2,1} & 1 \end{bmatrix} \tag{4.13}$$

and

$$\mathbf{E} = \begin{bmatrix} \sigma_5^2 & \times & \times & \times & \times \\ 0 & \sigma_4^2 & \times & \times & \times \\ 0 & 0 & \sigma_3^2 & \times & \times \\ 0 & 0 & 0 & \sigma_2^2 & \times \\ 0 & 0 & 0 & 0 & \sigma_1^2 \end{bmatrix}. \tag{4.14}$$

These matrices would change to

$$\mathbf{A}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ \alpha_{5,1} & 1 & 0 & 0 & 0 \\ \alpha_{5,2} & \alpha_{4,1} & 1 & 0 & 0 \\ \alpha_{5,3} & \alpha_{4,2} & \alpha_{3,1} & 1 & 0 \\ \alpha_{5,4} & \alpha_{4,3} & \alpha_{3,2} & 0 & 1 \end{bmatrix} \tag{4.15}$$

and

$$\mathbf{E} = \begin{bmatrix} \sigma_5^2 & \times & \times & \times & \times \\ 0 & \sigma_4^2 & \times & \times & \times \\ 0 & 0 & \sigma_3^2 & \times & \times \\ 0 & 0 & 0 & \sigma_2^2 & \sigma_{21}^2 \\ 0 & 0 & 0 & \sigma_{12}^2 & \sigma_1^2 \end{bmatrix}. \tag{4.16}$$

because of the joint observations. Notice that the lower right block in $\mathbf{A}_0$ has been changed to the 2×2 identity matrix while the lower right block $\mathbf{E}$ is changed to a 2×2 error covariance matrix which characterizes the joint prediction of x_1 and x_2. The equations for solving the Normal equations must be slightly modified to account for any joint observations. In particular, the covariance elements of the joint observations can be calculated from

$$\sigma_{ij}^2 = \mathcal{E}\{x_i[m]\varepsilon_j^*[m]\} = R_{x_i x_j}[0] + \sum_p \alpha_{i,p} R_{x_i x_{i-p}}[0] + \sum_{q=1}^{Q} \mathbf{a}_{i,q}^T \mathbf{r}_{x_i \mathbf{x}}[q], \tag{4.17}$$

where i and j are jointly observed, the summation over p does not include any joint observations and $\sigma_i^2 = \sigma_{ii}^2$. In addition, the orthogonality equations become

$$\begin{aligned} 0 &= \mathcal{E}\{x_j[m-l]\varepsilon_i^*[m]\} \\ &= R_{x_j x_i}[-l] + \sum_p \alpha_{i,p} R_{x_j x_{i-p}}[-l] + \sum_{q=1}^{Q} \mathbf{a}_{i,q}^T \mathbf{r}_{x_j \mathbf{x}}[q-l]. \end{aligned} \tag{4.18}$$

The explicit form of the multirate linear prediction problem can most easily be seen by using an example. Figure 4.3 shows a two-channel multirate system with decimation factors of $K_1 = 1$ and $K_2 = 2$. Channel 1 has two samples per period while channel 2 has one sample per period. In addition, observations from both channels occur simultaneously at every other system clock interval. Thus joint prediction of the channels must occur at those times. The block variables $x_i[m]$ used in this example have been labeled in Fig. 4.3. Using a filter order of $P = 1$ and assuming that the block of observations at $m = 0$ are available, the linear prediction equations for predicting $x_i[m]$ can be determined. For $x_1[1]$ the equations would be

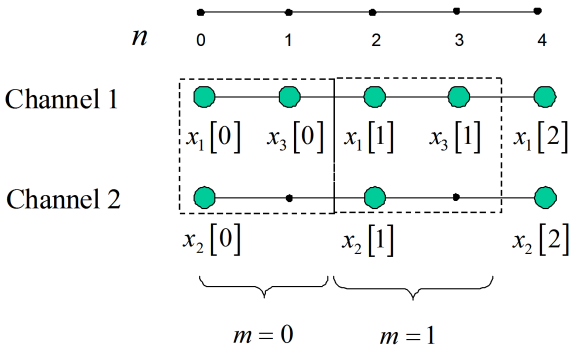


Figure 4.3. Illustration of Decimation for a Linear Prediction System.

$$\hat{x}_1[1] = -a^*_{1,1}x_3[0] - a^*_{1,2}x_2[0] - a^*_{1,3}x_1[0]$$

The error would then be $\varepsilon_1[1] = x_1[1] - \hat{x}_1[1]$ or

$$\varepsilon_1[1] = x_1[1] + a^*_{1,1}x_3[0] + a^*_{1,2}x_2[0] + a^*_{1,3}x_1[0]$$

Applying (4.17) with $i = j = 1$, the prediction error variance σ_1^2 would be

$$\sigma_1^2 = R_{x_1}[0] + a_{1,1}R_{x_1x_3}[1] + a_{1,2}R_{x_1x_2}[1] + a_{1,3}R_{x_1}[1] \tag{4.19}$$

and the prediction error covariance for observation $x_2[1]$ and $x_1[1]$ is

$$\sigma_{21}^2 = R_{x_2x_1}[0] + a_{1,1}R_{x_2x_3}[1] + a_{1,2}R_{x_2}[1] + a_{1,3}R_{x_2x_1}[1] \tag{4.20}$$

Applying (4.18), the orthogonality equations for the previous observations are

$$R_{x_3x_1}[-1] + \quad a_{1,1}R_{x_3}[0] + a_{1,2}R_{x_3x_2}[0] + a_{1,3}R_{x_3x_1}[0] = 0 \tag{4.21}$$

$$R_{x_2x_1}[-1] + a_{1,1}R_{x_2x_3}[0] + \quad a_{1,2}R_{x_2}[0] + a_{1,3}R_{x_2x_1}[0] = 0 \tag{4.22}$$

$$R_{x_1}[-1] + a_{1,1}R_{x_1x_3}[0] + a_{1,2}R_{x_1x_2}[0] + \quad a_{1,3}R_{x_1}[0] = 0 \tag{4.23}$$

In a similar fashion, the linear prediction equations for $x_2[1]$ can be written. The prediction equation, using (4.3), is

$$\hat{x}_2[1] = -a^*_{2,1}x_1[0] - a^*_{2,2}x_2[0] - a^*_{2,3}x_1[0]$$

with an error equation, from (4.5), of

$$\varepsilon_2[1] = x_2[1] + a^*_{2,1}x_3[0] + a^*_{2,2}x_2[0] + a^*_{2,3}x_1[0]$$

The prediction error variance and prediction error covariance equations, derived from (4.17), are

$$\sigma_2^2 = R_{x_2}[0] + a_{2,1}R_{x_2x_3}[1] + a_{2,2}R_{x_2}[1] + a_{2,3}R_{x_2x_1}[1] \tag{4.24}$$

$$\sigma_{12}^2 = R_{x_1x_2}[0] + a_{2,1}R_{x_1x_3}[1] + a_{2,2}R_{x_1x_2}[1] + a_{2,3}R_{x_1}[1] \tag{4.25}$$

The orthogonality equations for the previous observations, using (4.18), are

$$R_{x_3x_2}[-1] + \ a_{2,1}R_{x_3}[0] + a_{2,2}R_{x_3x_2}[0] + a_{2,3}R_{x_3x_1}[0] = 0 \tag{4.26}$$

$$R_{x_2}[-1] + a_{2,1}R_{x_2x_3}[0] + \ a_{2,2}R_{x_2}[0] + a_{2,3}R_{x_2x_1}[0] = 0 \tag{4.27}$$

$$R_{x_1x_2}[-1] + a_{2,1}R_{x_1x_3}[0] + a_{2,2}R_{x_1x_2}[0] + \ a_{2,3}R_{x_1}[0] = 0 \tag{4.28}$$

Finally, the equations for $x_3[1]$ can be written as

$$\hat{x}_3[1] = -\alpha^*_{3,1}x_2[1] - \alpha^*_{3,2}x_1[1] - a^*_{3,1}x_3[0] - a^*_{3,2}x_2[0] - a^*_{3,3}x_1[0]$$

$$\varepsilon_3[1] = x_3[1] + \alpha^*_{3,1}x_2[1] + \alpha^*_{3,2}x_1[1] + a^*_{3,1}x_3[0] + a^*_{3,2}x_2[0] + a^*_{3,3}x_1[0]$$

and

$$\sigma_3^2 = R_{x_3}[0] + \alpha_{3,1}R_{x_3x_2}[0] + \alpha_{3,2}R_{x_3x_1}[0] + a_{3,1}R_{x_3}[1] + a_{3,2}R_{x_3x_2}[1] + a_{3,3}R_{x_3x_1}[1] \tag{4.29}$$

There are no prediction covariance equations since there are no observations occurring at the same time as $x_3[1]$. In this case however, the number of orthogonality equations increases to five, since the error $\varepsilon_3[1]$ must be orthogonal to all five previous

observations (see Fig. 4.3) These orthogonality equations are:

$$R_{x_2x_3}[0]+ \quad \alpha_{3,1}R_{x_2}[0]+ \quad \alpha_{3,2}R_{x_2x_1}[0]+a_{3,1}R_{x_2x_3}[1]+ \quad a_{3,2}R_{x_2}[1]+a_{3,3}R_{x_2x_1}[1] = 0 \tag{4.30}$$

$$R_{x_1x_3}[0]+ \quad \alpha_{3,1}R_{x_1x_2}[0]+ \quad \alpha_{3,2}R_{x_1}[0]+a_{3,1}R_{x_1x_3}[1]+a_{3,2}R_{x_1x_2}[1]+ \quad a_{3,3}R_{x_1}[1] = 0 \tag{4.31}$$

$$R_{x_3}[-1]+\alpha_{3,1}R_{x_3x_2}[-1]+\alpha_{3,2}R_{x_3x_1}[-1]+ \quad a_{3,1}R_{x_3}[0]+a_{3,2}R_{x_3x_2}[0]+a_{3,3}R_{x_3x_1}[0] = 0 \tag{4.32}$$

$$R_{x_2x_3}[-1]+ \quad \alpha_{3,1}R_{x_2}[-1]+\alpha_{3,2}R_{x_2x_1}[-1]+a_{3,1}R_{x_2x_3}[0]+ \quad a_{3,2}R_{x_2}[0]+a_{3,3}R_{x_2x_1}[0] = 0 \tag{4.33}$$

$$R_{x_1x_3}[-1] + \alpha_{3,1}R_{x_1x_2}[-1] + \alpha_{3,2}R_{x_1}[-1] + a_{3,1}R_{x_1x_3}[0] + a_{3,2}R_{x_1x_2}[0] + a_{3,3}R_{x_1}[0] = 0 \tag{4.34}$$

By combining (4.19) through (4.34), the matrix form of the multichannel Normal equations (4.9) can be produced and written as

$$\begin{bmatrix} \mathbf{R}_0 & \vdots & \mathbf{R}_1 \\ \cdots & \cdots & \cdots \\ \mathbf{R}_{-1} & \vdots & \mathbf{R}_0 \end{bmatrix} \begin{bmatrix} \mathbf{A}_0 \\ \cdots \\ \mathbf{A}_1 \end{bmatrix} = \begin{bmatrix} \mathbf{E} \\ \cdots \\ \mathbf{0} \end{bmatrix},$$

where

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_0 & \vdots & \mathbf{R}_1 \\ \cdots & \cdots & \cdots \\ \mathbf{R}_{-1} & \vdots & \mathbf{R}_0 \end{bmatrix} \tag{4.35}$$

or, more explicitly,

$$\mathbf{R} = \left[\begin{array}{ccc:c:ccc} R_{x_3}[0] & R_{x_3x_2}[0] & R_{x_3x_1}[0] & \vdots & R_{x_3}[1] & R_{x_3x_2}[1] & R_{x_3x_1}[1] \\ R_{x_2x_3}[0] & R_{x_2}[0] & R_{x_2x_1}[0] & \vdots & R_{x_2x_3}[1] & R_{x_2}[1] & R_{x_2x_1}[1] \\ R_{x_1x_3}[0] & R_{x_1x_2}[0] & R_{x_1}[0] & \vdots & R_{x_1x_3}[1] & R_{x_1x_2}[1] & R_{x_1}[1] \\ \hdashline R_{x_3}[-1] & R_{x_3x_2}[-1] & R_{x_3x_1}[-1] & \vdots & R_{x_3}[0] & R_{x_3x_2}[0] & R_{x_3x_1}[0] \\ R_{x_2x_3}[-1] & R_{x_2}[-1] & R_{x_2x_1}[-1] & \vdots & R_{x_2x_3}[0] & R_{x_2}[0] & R_{x_2x_1}[0] \\ R_{x_1x_3}[-1] & R_{x_1x_2}[-1] & R_{x_1}[-1] & \vdots & R_{x_1x_3}[0] & R_{x_1x_2}[0] & R_{x_1}[0] \end{array}\right]$$

$$\mathbf{A} = \left[\begin{array}{c} \mathbf{A}_0 \\ \hdashline \mathbf{A}_1 \end{array}\right] = \left[\begin{array}{ccc} 1 & 0 & 0 \\ \alpha_{3,1} & 1 & 0 \\ \alpha_{3,2} & 0 & 1 \\ \hdashline a_{3,1} & a_{2,1} & a_{1,1} \\ a_{3,2} & a_{2,2} & a_{1,2} \\ a_{3,3} & a_{2,3} & a_{1,3} \end{array}\right] \tag{4.36}$$

and the error matrix is

$$\left[\begin{array}{c} \mathbf{E} \\ \hdashline \mathbf{0} \end{array}\right] = \left[\begin{array}{ccc} \sigma_3^2 & \times & \times \\ 0 & \sigma_2^2 & \sigma_{21}^2 \\ 0 & \sigma_{12}^2 & \sigma_1^2 \\ \hdashline 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right] \tag{4.37}$$

C. AN EFFICIENT SOLUTION TO THE MULTIRATE NORMAL EQUATIONS

In the previous section, the Normal equations for the multirate linear prediction problem were developed element by element. Although these equations can be solved directly, it is desirable to develop a method for calculating the linear prediction coefficients and error variances more efficiently. Efficient recursive methods for solving these equations have been developed for both the single-channel and multichannel cases. In the single-channel case the method is called the Levinson (or Levinson-Durbin) algorithm while in the multichannel case it is known commonly as the multichannel Levinson algorithm, or in some literature it is called the Levinson-Wiggins-Robinson (LWR) algorithm [Ref. 64]. In the following, a similar procedure is developed for the multirate case.

The equations to be solved are

$$\begin{bmatrix} \mathbf{R}_0 & \mathbf{R}_1 & \cdots & \mathbf{R}_Q \\ \mathbf{R}_{-1} & \mathbf{R}_0 & \cdots & \mathbf{R}_{Q-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{-Q} & \mathbf{R}_{1-Q} & \cdots & \mathbf{R}_0 \end{bmatrix} \begin{bmatrix} \mathbf{A}_0 \\ \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_Q \end{bmatrix} = \begin{bmatrix} \mathbf{E} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}. \tag{4.38}$$

These are the multirate Normal equations of (4.9), repeated here for convenience.

To begin the solution to (4.38), consider the problem of predicting an entire block of data at once. This is identical to the multichannel problem. The Normal equations for this problem are of the form (4.2) and can be written as

$$\begin{bmatrix} \mathbf{R}_0 & \mathbf{R}_1 & \cdots & \mathbf{R}_Q \\ \mathbf{R}_{-1} & \mathbf{R}_0 & \cdots & \mathbf{R}_{Q-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{-Q} & \mathbf{R}_{1-Q} & \cdots & \mathbf{R}_0 \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ \mathbf{A}_1' \\ \vdots \\ \mathbf{A}_Q' \end{bmatrix} = \begin{bmatrix} \mathbf{E}' \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}. \tag{4.39}$$

Observe that the correlation matrix in (4.39) is identical to that in (4.38) although the variables $\mathbf{A}_i^{'}$ and $\mathbf{E}^{'}$ are not. If (4.39) is right multiplied with $\mathbf{A}_0$ however, and (4.38) compared to (4.39), it can be seen that the following equalities must hold:

$$\mathbf{A}_i = \mathbf{A}_i^{'}\mathbf{A}_0, \qquad i = 1, \ldots, Q \tag{4.40}$$

$$\mathbf{E} = \mathbf{E}^{'}\mathbf{A}_0. \tag{4.41}$$

Therefore, if solutions to (4.39) and the matrix $\mathbf{A}_0$ can be found, then the desired multichannel parameters $\mathbf{A}_i$ (for $i = 1, 2, \ldots, Q$) and $\mathbf{E}$ can be found from (4.40) and (4.41).

A computationally efficient solution to (4.39) can be found using the LWR algorithm (see Appendix D). This solution produces the matrices $\mathbf{A}_i^{'}$, where $i = 1, 2, \ldots, Q$ and $\mathbf{E}^{'}$. To find $\mathbf{A}_0$ from (4.41) the *forms* of $\mathbf{E}$ and $\mathbf{A}_0$ must be considered. In particular, $\mathbf{E}$ is upper triangular, e.g., (4.14), when there are no simultaneous observations, and *block* upper triangular, e.g., (4.16), when some observations occur simultaneously. Also, $\mathbf{A}_0$ is lower triangular with unit diagonal, e.g., (4.13), when there are no simultaneous observations, and block lower triangular in general, e.g., (4.15). Consider the specific case shown in Fig. 4.2. By writing (4.41) as $\mathbf{E}^{'} = \mathbf{E}\mathbf{A}_0^{-1}$, the matrices have the form

$$\begin{bmatrix} \sigma_5^{\prime 2} & \sigma_{54}^{\prime 2} & \sigma_{53}^{\prime 2} & \sigma_{52}^{\prime 2} & \sigma_{51}^{\prime 2} \\ \sigma_{45}^{\prime 2} & \sigma_4^{\prime 2} & \sigma_{43}^{\prime 2} & \sigma_{42}^{\prime 2} & \sigma_{41}^{\prime 2} \\ \sigma_{35}^{\prime 2} & \sigma_{34}^{\prime 2} & \sigma_3^{\prime 2} & \sigma_{32}^{\prime 2} & \sigma_{31}^{\prime 2} \\ \sigma_{25}^{\prime 2} & \sigma_{24}^{\prime 2} & \sigma_{23}^{\prime 2} & \sigma_2^{\prime 2} & \sigma_{21}^{\prime 2} \\ \sigma_{15}^{\prime 2} & \sigma_{14}^{\prime 2} & \sigma_{13}^{\prime 2} & \sigma_{12}^{\prime 2} & \sigma_1^{\prime 2} \end{bmatrix} = \begin{bmatrix} \sigma_5^2 & \times & \times & \times & \times \\ 0 & \sigma_4^2 & \times & \times & \times \\ 0 & 0 & \sigma_3^2 & \times & \times \\ 0 & 0 & 0 & \sigma_2^2 & \sigma_{21}^2 \\ 0 & 0 & 0 & \sigma_{12}^2 & \sigma_1^2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ \alpha_{5,1} & 1 & 0 & 0 & 0 \\ \alpha_{5,2} & \alpha_{4,1} & 1 & 0 & 0 \\ \alpha_{5,3} & \alpha_{4,2} & \alpha_{3,1} & 1 & 0 \\ \alpha_{5,4} & \alpha_{4,3} & \alpha_{3,2} & 0 & 1 \end{bmatrix}^{-1}. \tag{4.42}$$

It can be seen that (4.42) is similar in form to $\mathbf{M} = \mathbf{UL}$, where $\mathbf{U}$ and $\mathbf{L}$ are upper and lower triangular matrices, respectively. Thus if $\mathbf{E}^{'}$ can be factored into $\mathbf{E}$ and $\mathbf{A}_0^{-1}$, then the solution to the multirate Normal equations has been found.

Most linear algebra books (e.g., [Ref. 65]) contain LU factorization algorithms for solving the decomposition of a matrix into a lower-upper triangular scheme (i. e.,

$\mathbf{M} = \mathbf{LU}$). For the purpose of this dissertation, however, it is desired to factor $\mathbf{E}'$ into an upper-lower scheme. In addition, this factorization must be general enough to allow for block elimination in the algorithm (this accounts for joint observations in the multirate system). Thus a generalized UL factorization algorithm has been written and provided in Appendix E. Using this generalized UL factorization algorithm to find $\mathbf{A}_0$, the multirate Normal equations can be solved. A significant advantage of decomposing $\mathbf{E}'$ into $\mathbf{E}$ and $\mathbf{A}_0^{-1}$ is that the only matrix that requires inversion is $\mathbf{A}_0^{-1}$. Since, $\mathbf{A}_0^{-1}$ is lower triangular, $\mathbf{A}_0$ can be found efficiently using forward elimination which does not explicitly perform a full matrix inversion [Ref. 66].

The steps needed to solve for the multirate Normal equations are summarized in Table 4.1.

Table 4.1. Steps in Solving the Multirate Normal Equations

1.	Solve (4.39) using the multichannel Levinson recursion to find $\mathbf{E}', \mathbf{A}_1', \ldots, \mathbf{A}_Q'$
2.	Factor $\mathbf{E}'$ using the generalized UL factorization algorithm to obtain $\mathbf{E}$ and $\mathbf{A}_0^{-1}$
3.	Invert $\mathbf{A}_0^{-1}$ to obtain $\mathbf{A}_0$
4.	Compute $\mathbf{A}_i = \mathbf{A}_i'\mathbf{A}_0$ for $i = 1, \ldots, Q$

D. SUMMARY

In this chapter, the multirate linear prediction equations were developed. Also, an efficient method for calculating the linear prediction coefficients using the multichannel Levinson recursion and LU factorization was propsed. These techniques for calculating the linear prediction coefficients are used in the next chapter on multirate sequential classification.

V. CLASSIFICATION

The problem of distinguishing between two or more different types of signals (arising from different sources) is known as classification. The classification is said to be "sequential" when the signal to be classified is observed one discrete sample at a time. After a new sample is observed, an attempt is made to classify the signal using all of the samples available up to that time. If a classification cannot be made, then another sample is taken and the procedure continues. The goal is to classify the signal with a prescribed probability of error using as few samples as possible. The general theory of sequential classification was originally and extensively developed by Wald in 1947 [Ref. 20]. It has since been applied specifically to statistical signal processing and described for the single-channel and multichannel cases (e.g., [Ref. 22] and [Ref. 21].) The focus of this chapter is to demonstrate the feasibility of developing an algorithm that allows for multirate observations to be used in a sequential classifier.

A. SEQUENTIAL CLASSIFICATION

Before developing the sequential classifier for the multirate system, it is necessary to provide a brief description of the sequential classification process. Conventional methods of classification usually involve a fixed block of data. This method is known as classifying with a fixed sample size. Statistically optimal methods then involve developing the ratio of likelihood functions for the two classes and comparing that likelihood ratio to a fixed threshold. This threshold is chosen to optimize some criterion, such as total probability of error (Bayes test) or maximizing the probability of correctly choosing one class while holding the probability of error on the other class fixed (Neyman-Pearson test) [Ref. 22, 67, 68]. Wald developed a method, called the *sequential probability ratio test* (SPRT), which allowed one to successively add samples to data set being used to perform the classification. If classification is not

possible at that time, another sample is added. This process is repeated until a classification can be made with some fixed desired probability of error. For an extensive discussion on sequential analysis, see [Ref. 20].

To begin the discussion of sequential classification, assume that the signal of interest, x, belongs to one of only two classes, class 0 or class 1. Let H_0 be the hypothesis that the signal belongs to class 0 and H_1 the hypothesis that the signal belongs to class 1. Let the vector of observations be denoted by

$$\mathbf{x}_n = \begin{bmatrix} \underline{x}_0 \\ \vdots \\ \underline{x}_n \end{bmatrix}, \tag{5.1}$$

where $\underline{x}_j$ is the set of observations of x available at time j, and let $f(\mathbf{x}_n|H_i)$ be the probability density function for $\mathbf{x}_n$ given that the observations are from class i. In the Wald SPRT, the *likelihood ratio* is defined as

$$l_n(\mathbf{x}_n) = \frac{f(\mathbf{x}_n|H_1)}{f(\mathbf{x}_n|H_0)} \tag{5.2}$$

and two positive constants or *thresholds* A and B (with $A > B$) are chosen to classify the signal $\mathbf{x}_n$ as either class 0 or class 1 with some desired probabilities of error. The test performed is as follows: if

$$\frac{f(\mathbf{x}_n|H_1)}{f(\mathbf{x}_n|H_0)} \geq A \tag{5.3}$$

then $\mathbf{x}_n$ is classified as class 1; while if

$$\frac{f(\mathbf{x}_n|H_1)}{f(\mathbf{x}_n|H_0)} \leq B \tag{5.4}$$

then $\mathbf{x}_n$ is classified as class 0. If the ratio of the conditional probabilities falls between A and B, that is, if

$$B < \frac{f(\mathbf{x}_n|H_1)}{f(\mathbf{x}_n|H_0)} < A, \tag{5.5}$$

then another sample is taken and the process is repeated with

$$\mathbf{x}_{n+1} = \begin{bmatrix} \underline{x}_1 \\ \vdots \\ \underline{x}_n \\ \underline{x}_{n+1} \end{bmatrix}.$$

Proper selection of the values for A and B is discussed below.

In many cases, it is more convenient to work with the logarithm of the likelihood ratio rather than the likelihood ratio itself. Since the logarithm is a strictly monotonic function, this does not change the essential nature of the SPRT. In the particular case that follows, it is most convenient to deal with the *negative* logarithm and define

$$h_n(\mathbf{x}_n) = -2 \ln(l_n(\mathbf{x}_n)) \tag{5.6}$$

with

$$\tau_A = -2 \ln(A) \tag{5.7}$$

$$\tau_B = -2 \ln(B) \tag{5.8}$$

The inequality then reverses and the SPRT decision becomes

$$\boxed{\begin{array}{lll} h_n(\mathbf{x}_n) & \leq \tau_A & \text{choose } H_1 \\ & \geq \tau_B & \text{choose } H_0 \end{array}}, \tag{5.9}$$

where

$$h_n(\mathbf{x}_n) = -2 \ln \frac{f_{\mathbf{x}_n|H_1}(\mathbf{x}_n|H_1)}{f_{\mathbf{x}_n|H_0}(\mathbf{x}_n|H_0)}. \tag{5.10}$$

Otherwise take another observation.

The choices for the thresholds A and B are related to the probabilities of error for class 0 and class 1. Using the method described in [Ref. 22, 21], the relations between thresholds and probabilities of error can be derived as follows. Let ϵ_0 be the probability of a decision error given the data is from class 0 and ϵ_1 be the probability

of error given the data is from class 1. Assume that the upper threshold is crossed after n observations; that is, at time n the ratio $l_n(\mathbf{x}_n)$ defined by (5.2) is greater than but approximately equal to A:

$$l_n(\mathbf{x}_n) \gtrsim A \tag{5.11}$$

Substituting (5.2) into (5.11) and rearranging yields

$$f(\mathbf{x}_n|H_1) \gtrsim A\, f(\mathbf{x}_n|H_0). \tag{5.12}$$

The condition (5.12) defines a region of the $\mathbf{x}_n$ space (R_1) where the data is classified as class 1 (H_1). Integrating both sides over this region yields

$$\int_{R_1} f(\mathbf{x}_n|H_1)d\mathbf{x}_n \gtrsim A \int_{R_1} f(\mathbf{x}_n|H_0)d\mathbf{x}_n. \tag{5.13}$$

Now, since R_1 is taken to be the region of measurement space where $\mathbf{x}_n$ is classified as class 1, the left hand side of (5.13) is the probability of correctly classifying class 1, namely

$$\int_{R_1} f(\mathbf{x}_n|H_1)d\mathbf{x}_n = 1 - \epsilon_1. \tag{5.14}$$

The integral on the right hand side of (5.13) is then the probability of incorrectly classifying class 0,

$$\int_{R_1} f(\mathbf{x}_n|H_0)d\mathbf{x}_n = \epsilon_0 \tag{5.15}$$

Substituting (5.14) and (5.15) into (5.13) yields

$$1 - \epsilon_1 \gtrsim A\epsilon_0 \tag{5.16}$$

or

$$A = \frac{1 - \epsilon_1}{\epsilon_0} \tag{5.17}$$

where the inequality has been replaced by an equal sign since the threshold A is most restrictive at that value. Similarly, assuming

$$l_n(\mathbf{x}_n) \lesssim B \tag{5.18}$$

and integrating over the region where $\mathbf{x}_n$ is classified as class 0 yields

$$B = \frac{\epsilon_1}{1 - \epsilon_0} \tag{5.19}$$

Equations (5.17) and (5.19) provide the necessary relations to allow one to choose appropriate values of thresholds A and B based on desired probabilities of error.

B. ALGORITHM DEVELOPMENT

When the observations in the SPRT are jointly Gaussian for both classes, then a convenient recursive classification algorithm can be developed that involves linear prediction. This recursive form has been developed in [Ref. 22, 21] for the case of a single observation sequence, which is scalar-valued or vector-valued (the multichannel case). Here, the method is extended to the case of multiple channels sampled at different rates. Thus, at any given epoch on the system time scale, there may be one or more simultaneous observations available from the various channels or possibly no observations. The algorithm to be described takes this more general situation into account.

The development of the recursive form of the quadratic classifier for the multirate system closely follows the development for the single-channel and multichannel cases developed in [Ref. 22] and [Ref. 21]. It is necessary, however, to make modifications to some of the parameters, to account for the periodic nature of the multirate system. Thus, some parameters that are constant in the single-channel and multichannel cases become periodically time-varying variables in the multirate case.

To start the development, assume that the signals associated with both classes are jointly Gaussian and that the observations for all channels are collected and ordered such that $\mathbf{x}_n$ contains all observations from time 0 to n. In addition, the observation vector (5.1) can be written recursively as

$$\mathbf{x}_n = \begin{bmatrix} \mathbf{x}_{n-1} \\ \underline{x}_n^{(k)} \end{bmatrix}, \tag{5.20}$$

where the superscript k is used to indicate that the size of the observation vector $\underline{x}_n^{(k)}$ is a periodically time-varying parameter since the number of observed channels at any time step varies periodically. If the size of $\underline{x}_n^{(k)}$ (the number of observations at time n) is defined as $S_n^{(k)}$ and the total number of observations from time 0 to $n-1$ is defined as N_{n-1}, then the total number of observations at time n can be expressed as

$$N_n = N_{n-1} + S_n^{(k)}. \tag{5.21}$$

As an example, consider the two-channel system where the first channel is decimated by a factor of 2 and the second channel is decimated by a factor of 3, as shown in Fig. 5.1. The system period as defined in Chapter II is $K = 6$. Thus, since $n \equiv k \pmod{K}$, the system phase, k, varies periodically between 0 and 5.

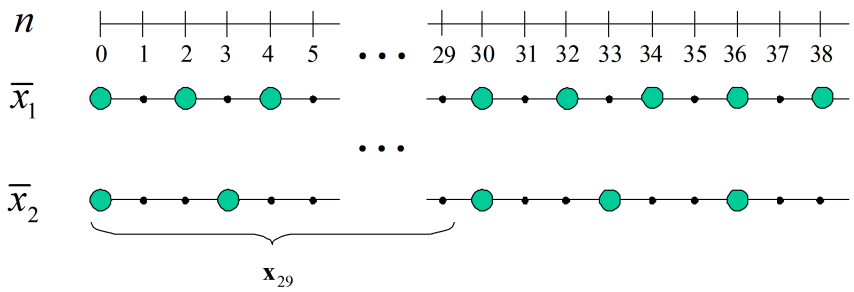


Figure 5.1. Two-channel Multirate System.

This can be illustrated more specifically as follows (refer to Fig. 5.1). Assuming that observations through $n = 29$ have been collected, the associated observation vectors for time steps $n = 30, 31, \ldots, 35$ would be denoted by

$$\mathbf{x}_{30} = \begin{bmatrix} \mathbf{x}_{29} \\ \underline{x}_{30}^{(0)} \end{bmatrix} \qquad \mathbf{x}_{31} = \begin{bmatrix} \mathbf{x}_{30} \\ \underline{x}_{31}^{(1)} \end{bmatrix} \qquad \mathbf{x}_{32} = \begin{bmatrix} \mathbf{x}_{31} \\ \underline{x}_{32}^{(2)} \end{bmatrix}$$

$$\mathbf{x}_{33} = \begin{bmatrix} \mathbf{x}_{32} \\ \underline{x}_{33}^{(3)} \end{bmatrix} \qquad \mathbf{x}_{34} = \begin{bmatrix} \mathbf{x}_{33} \\ \underline{x}_{34}^{(4)} \end{bmatrix} \qquad \mathbf{x}_{35} = \begin{bmatrix} \mathbf{x}_{34} \\ \underline{x}_{35}^{(5)} \end{bmatrix}$$

where the vectors of observations at these times are

$$\underline{x}_{30}^{(0)} = \begin{bmatrix} \bar{x}_1[30] \\ \bar{x}_2[30] \end{bmatrix} \qquad \underline{x}_{31}^{(1)} = \Big[\text{empty}\Big] \qquad \underline{x}_{32}^{(2)} = \bar{x}_1[32]$$

$$\underline{x}_{33}^{(3)} = \bar{x}_2[33] \qquad \underline{x}_{34}^{(4)} = \bar{x}_1[34] \qquad \underline{x}_{35}^{(5)} = \Big[\text{empty}\Big]$$

The corresponding number of observations, $S_n^{(k)}$, at these times are

$$S_{30}^{(0)} = 2 \qquad S_{31}^{(1)} = 0 \qquad S_{32}^{(2)} = 1$$

$$S_{33}^{(3)} = 1 \qquad S_{34}^{(4)} = 1 \qquad S_{35}^{(5)} = 0$$

For any given time 'n', the probability density function for the collection of observations is given by the multirate Gaussian form

$$f_{\mathbf{x}_n}(\mathbf{x}_n) = \frac{1}{(2\pi)^{N_n/2}|\mathbf{C}_n|^{1/2}} \mathrm{e}^{-\frac{1}{2}(\mathbf{x}_n - \mathbf{m}_n)^T \mathbf{C}_n^{-1} (\mathbf{x}_n - \mathbf{m}_n)}, \tag{5.22}$$

where $\mathbf{m}_n = \mathcal{E}\{\mathbf{x}_n\}$ is the mean vector and $\mathbf{C}_n = \mathcal{E}\{(\mathbf{x}_n - \mathbf{m}_n)(\mathbf{x}_n - \mathbf{m}_n)^T\}$ is the covariance matrix for the observations. The mean vector and covariance matrix can then be written as

$$\mathbf{m}_n = \begin{bmatrix} \mathbf{m}_{n-1} \\ \underline{m}_n^{(k)} \end{bmatrix} \tag{5.23}$$

and

$$\mathbf{C}_n = \begin{bmatrix} \mathbf{C}_{n-1} & \mathbf{R}_n^{(k)} \\ \mathbf{R}_n^{(k)T} & \mathbf{\Sigma}_n^{(k)} \end{bmatrix} \tag{5.24}$$

where the partitioning corresponds to the partitioning defined in (5.20). By using a formula for matrix partitioning [Ref. 69], the inverse covariance matrix, $\mathbf{C}_n^{-1}$, can be represented in terms of the inverse covariance matrix, $\mathbf{C}_{n-1}^{-1}$, as

$$\mathbf{C}_n^{-1} = \begin{bmatrix} \mathbf{I} & -\mathbf{G}_n^{(k)} \\ \mathbf{0}^T & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{C}_{n-1}^{-1} & \mathbf{0} \\ \mathbf{0}^T & \mathbf{E}_n^{(k)^{-1}} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{G}_n^{(k)T} & \mathbf{I} \end{bmatrix} \tag{5.25}$$

or equivalently as

$$\mathbf{C}_n^{-1} = \begin{bmatrix} \mathbf{C}_{n-1}^{-1} & \mathbf{0} \\ \mathbf{0}^T & \mathbf{0} \end{bmatrix} + \begin{bmatrix} -\mathbf{G}_n^{(k)} \\ \mathbf{I} \end{bmatrix} \mathbf{E}_n^{(k)^{-1}} \begin{bmatrix} -\mathbf{G}_n^{(k)^T} & \mathbf{I} \end{bmatrix} \tag{5.26}$$

where $\mathbf{G}_n^{(k)}$ and $\mathbf{E}_n^{(k)}$ are defined as

$$\mathbf{G}_n^{(k)} = \mathbf{C}_{n-1}^{-1}\mathbf{R}_n^{(k)} \tag{5.27}$$

$$\mathbf{E}_n^{(k)} = \mathbf{\Sigma}_n^{(k)} - \mathbf{R}_n^{(k)^T}\mathbf{C}_{n-1}^{-1}\mathbf{R}_n^{(k)}. \tag{5.28}$$

Since the determinants of the upper and lower triangular matrices of (5.25) are unity, the determinant of $\mathbf{C}_n^{-1}$ is given by

$$|\mathbf{C}_n^{-1}| = |\mathbf{C}_{n-1}^{-1}||\mathbf{E}_n^{(k)^{-1}}|$$

or

$$|\mathbf{C}_n| = |\mathbf{C}_{n-1}||\mathbf{E}_n^{(k)}|. \tag{5.29}$$

The partitionings of (5.1) and (5.23) through (5.29) provide all the necessary relations for the recursive computation of the density function (5.22) at each time step.

Now that the probability density function for the multirate system has been described, the probability density function for a given class within the multirate system can be discussed. Consider observations from a known class i with a corresponding mean vector

$$\mathbf{m}_n^{(i)} = \begin{bmatrix} \mathbf{m}_{n-1}^{(i)} \\ \underline{m}_n^{(i)(k)} \end{bmatrix} \tag{5.30}$$

and covariance matrix

$$\mathbf{C}_n^{(i)} = \begin{bmatrix} \mathbf{C}_{n-1}^{(i)} & \mathbf{R}_n^{(i)(k)} \\ \mathbf{R}_n^{(i)(k)^T} & \mathbf{\Sigma}_n^{(i)(k)} \end{bmatrix}, \tag{5.31}$$

where the superscript (i) denotes association with class i and the partitioning corresponds to the partitioning defined in (5.23) and (5.24). The vector of observations with the mean removed can be denoted as

$$\mathbf{x}_n^{(i)} = \mathbf{x}_n - \mathbf{m}_n^{(i)} \tag{5.32}$$

or

$$\mathbf{x}_n^{(i)} = \begin{bmatrix} \mathbf{x}_{n-1}^{(i)} \\ \underline{x}_n^{(i)(k)} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_{n-1} - \mathbf{m}_{n-1}^{(i)} \\ \underline{x}_n^{(k)} - \underline{m}_n^{(i)(k)} \end{bmatrix}. \tag{5.33}$$

By using the form (5.22) and substituting the relations (5.21), (5.32), (5.33), (5.26) and (5.29), the probability density function for class i at time n can be written as

$$f_{\mathbf{x}_n|H_i}(\mathbf{x}_n|H_i) = \frac{1}{(2\pi)^{(N_n)/2}|\mathbf{C}_n^{(i)}|^{1/2}} \mathrm{e}^{-\frac{1}{2}\left(\mathbf{x}_n - \mathbf{m}_n^{(i)}\right)^T \mathbf{C}_n^{(i)^{-1}} \left(\mathbf{x}_n - \mathbf{m}_n^{(i)}\right)}$$

$$= \frac{1}{(2\pi)^{N_{n-1}/2}|\mathbf{C}_{n-1}^{(i)}|^{1/2}} \mathrm{e}^{-\frac{1}{2}\left(\mathbf{x}_{n-1}^{(i)T} \mathbf{C}_{n-1}^{(i)^{-1}} \mathbf{x}_{n-1}^{(i)}\right)} \times \frac{1}{(2\pi)^{S_n^{(k)}/2}|\mathbf{E}_n^{(i)(k)}|^{1/2}}$$

$$\times \mathrm{e}^{-\frac{1}{2}\left(\underline{x}_n^{(i)(k)} - \mathbf{G}_n^{(k)^T} \mathbf{x}_{n-1}^{(i)}\right)^T \mathbf{E}_n^{(i)(k)^{-1}} \left(\underline{x}_n^{(i)(k)} - \mathbf{G}_n^{(k)^T} \mathbf{x}_{n-1}^{(i)}\right)}. \tag{5.34}$$

The first term on the right represents $f_{\mathbf{x}_{n-1}|H_i}$ while the second term on the right represents a conditional probability density function $f_{\underline{x}_n|\mathbf{x}_{n-1}H_i}$. In other words, (5.34) can be written as

$$f_{\mathbf{x}_n|H_i} = f_{\mathbf{x}_{n-1}|H_i} \cdot f_{\underline{x}_n|\mathbf{x}_{n-1}H_i} \qquad i = 1, 2 \tag{5.35}$$

where

$$f_{\mathbf{x}_{n-1}|H_i} = \frac{1}{(2\pi)^{N_{n-1}/2}|\mathbf{C}_{n-1}^{(i)}|^{1/2}} \mathrm{e}^{-\frac{1}{2}\left(\mathbf{x}_{n-1}^{(i)T} \mathbf{C}_{n-1}^{(i)^{-1}} \mathbf{x}_{n-1}^{(i)}\right)} \tag{5.36}$$

and

$$f_{\underline{x}_n|\mathbf{x}_{n-1}H_i} = \frac{1}{(2\pi)^{S_n^{(k)}/2}|\mathbf{E}_n^{(i)(k)}|^{1/2}} \mathrm{e}^{-\frac{1}{2}\left(\underline{x}_n^{(i)(k)} - \mathbf{G}_n^{(k)^T} \mathbf{x}_{n-1}^{(i)}\right)^T \mathbf{E}_n^{(i)(k)^{-1}} \left(\underline{x}_n^{(i)(k)} - \mathbf{G}_n^{(k)^T} \mathbf{x}_{n-1}^{(i)}\right)}. \tag{5.37}$$

Using these results, the statistic $h_n(\mathbf{x}_n)$ defined in (5.10) becomes

$$h_n(\mathbf{x}_n) = -2\ln\left(\frac{f_{\mathbf{x}_{n-1}|H_1}(\mathbf{x}_{n-1}|H_1)}{f_{\mathbf{x}_{n-1}|H_0}(\mathbf{x}_{n-1}|H_0)} \cdot \frac{f_{\underline{x}_n|\mathbf{x}_{n-1}H_1}}{f_{\underline{x}_n|\mathbf{x}_{n-1}H_0}}\right) \tag{5.38}$$

$$= -2\ln\frac{f_{\mathbf{x}_{n-1}|H_1}(\mathbf{x}_{n-1}|H_1)}{f_{\mathbf{x}_{n-1}|H_0}(\mathbf{x}_{n-1}|H_0)} - 2\ln\frac{f_{\underline{x}_n|\mathbf{x}_{n-1}H_1}}{f_{\underline{x}_n|\mathbf{x}_{n-1}H_0}}$$

This can be written as

$$h_n(\mathbf{x}_n) = h_{n-1}(\mathbf{x}_{n-1}) + \Delta h_n(\underline{x}_n^{(k)}|\mathbf{x}_{n-1}), \tag{5.39}$$

where

$$\Delta h_n(\underline{x}_n^{(k)}|\mathbf{x}_{n-1}) = -2\ln \frac{f_{\underline{x}_n|\mathbf{x}_{n-1}H_1}}{f_{\underline{x}_n|\mathbf{x}_{n-1}H_0}}. \tag{5.40}$$

Substituting (5.37) into (5.40) results in the expression

$$\Delta h_n(\underline{x}_n^{(k)}|\mathbf{x}_{n-1}) = \frac{(\underline{x}_n^{(1)} - \mathbf{G}_n^{(k)(1)^T}\mathbf{x}_{n-1}^{(1)})^2}{|\mathbf{E}_n^{(k)(1)}|} - \frac{(\underline{x}_n^{(0)} - \mathbf{G}_n^{(k)(0)^T}\mathbf{x}_{n-1}^{(0)})^2}{|\mathbf{E}_n^{(k)(0)}|} + \ln \frac{|\mathbf{E}_n^{(k)(1)}|}{|\mathbf{E}_n^{(k)(0)}|}. \tag{5.41}$$

Since the first observation vector is labeled $\mathbf{x}_0$ and the statistic $h_0(\mathbf{x}_0) = h_{-1}(\mathbf{x}_{-1}) + \Delta h_0(\underline{x}_0^{(0)}|\mathbf{x}_{-1})$, the SPRT is first initialized with $x_{-1} = 0$ and the statistic $h_{-1}(\mathbf{x}_{-1})$ is empty. Thereafter, h_n is computed from (5.39) and (5.41) recursively. This statistic is used in the decision rule (5.9) to perform the classification.

C. CLASSIFICATION USING LINEAR PREDICTION

The equations for the multirate sequential classifier were developed assuming that the observations were jointly Gaussian. The sequential classifier, however, can also be interpreted in terms of linear prediction. By rearranging (5.27) and (5.28) into

$$\mathbf{R}_n^{(k)} - \mathbf{C}_{n-1}\mathbf{G}_n^{(k)} = \mathbf{0} \tag{5.42}$$

$$\mathbf{E}_n^{(k)} = \mathbf{\Sigma}_n^{(k)} - \mathbf{R}_n^{(k)^T}\mathbf{G}_n^{(k)}, \tag{5.43}$$

these two equations can be combined in matrix form to write

$$\begin{bmatrix} \mathbf{\Sigma}_n^{(k)} & \mathbf{R}_n^{(k)^T} \\ \mathbf{R}_n^{(k)} & \mathbf{C}_{n-1} \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ -\mathbf{G}_n^{(k)} \end{bmatrix} = \begin{bmatrix} \mathbf{E}_n^{(k)} \\ \mathbf{0} \end{bmatrix}. \tag{5.44}$$

These equations bear a striking similarity to the Normal equations of linear prediction (see (4.2)). Indeed, the matrix $-\mathbf{G}_n^{(k)}$ can be interpreted as the prediction filter coefficients for predicting the new observations, and $\mathbf{E}_n^{(k)}$ is the prediction error

covariance matrix. More explicitly $-\mathbf{G}_n^{(k)}$ and $\mathbf{E}_n^{(k)}$ relate to the parameters in (4.9) by

$$\begin{bmatrix} \mathbf{A}_0 \\ \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_Q \end{bmatrix} = \left[\begin{array}{c|c|c|c} \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} \\ & \mathbf{I} & \dots & \mathbf{0} \\ -\mathbf{G}_n^{(K-1)} & & \ddots & \mathbf{0} \\ & -\mathbf{G}_n^{(K-2)} & \dots & \mathbf{I} \\ & & \dots & -\mathbf{G}_n^{(0)} \end{array}\right] \tag{5.45}$$

and

$$\mathbf{E} = \begin{bmatrix} \mathbf{E}_n^{(K-1)} & \times & \dots & \times \\ \mathbf{0} & \mathbf{E}_n^{(K-2)} & \dots & \times \\ \vdots & \vdots & \ddots & \times \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{E}_n^{(0)} \end{bmatrix}. \tag{5.46}$$

The block form in (5.45) and (5.46) may be a little misleading. In most cases, the width of the blocks is just one column and the "I" becomes a "1." The matrices in (5.45) and (5.46) have been represented in block form to allow for possible multiple observations at each time step. The parameters $-\mathbf{G}_n^{(k)}$ and $\mathbf{E}_n^{(k)}$ thus represent the periodically time-varying parameters used to predict each new observation (or set of observations) within a block.

When a sufficiently high order Q is chosen for the prediction, the parameters $\mathbf{G}_n^{(k)}$ and $\mathbf{E}_n^{(k)}$ do not depend explicitly on the time variable "n" but only on the phase as indicated by the superscript k.

D. ALGORITHM SUMMARY

This section summarizes the multirate sequential classifier algorithm for ease of implementation. The recursive classification algorithm can be divided into two parts: 1) training and 2) testing.

In the first part, the system is trained to distinguish between two classes using a priori information. Using known class training data, the classifier parameters (mean

vectors, prediction coefficient matrices and prediction error variance matrices) for each class are estimated. In addition, the thresholds used in classification are determined and adjusted experimentally. These parameters are summarized in Table 5.1.

Table 5.1. Summary of Classifier Training Parameters

Class i Parameters:	mean vectors, $\underline{m}_n^{(i)(k)} = \mathcal{E}_i[\underline{x}_n^{(k)}]$
$i = 0, 1$	Prediction coefficients, $\mathbf{G}_n^{(i)(k)}$
	Prediction Error Variances, $\mathbf{E}_n^{(i)(k)}$
Classification Thresholds:	A (Class 1) and B (Class 0)

In the second part new (test) data is presented to the classifier. The associated class errors ($\underline{\epsilon}_n^{(1)(k)^T}$ and $\underline{\epsilon}_n^{(0)(k)^T}$) for the given signal, $\mathbf{x}_n$, are calculated and used along with the class prediction error variances to calculate the SPRT statistic, h_n, for comparison with the class thresholds. When the SPRT statistic crosses a class threshold, the classifier declares the classification; otherwise, the process is repeated. These steps are shown in Table 5.2.

Table 5.2. Steps in Sequential Classification

1. $\mathbf{x}_n^{(i)} = \begin{bmatrix} \mathbf{x}_{n-1}^{(i)} \\ \underline{x}_n^{(k)} - \underline{m}_n^{(i)(k)} \end{bmatrix} \qquad i = 0, 1$

2. $\underline{\epsilon}_n^{(i)(k)} = \begin{bmatrix} -\mathbf{G}_n^{(i)(k)} \\ \mathbf{I} \end{bmatrix}^T \mathbf{x}_n^{(i)} \qquad i = 0, 1$

3. $\Delta h_n = \underline{\epsilon}_n^{(1)(k)^T} \mathbf{E}_n^{(1)(k)^{-1}} \underline{\epsilon}_n^{(1)(k)} - \underline{\epsilon}_n^{(0)(k)^T} \mathbf{E}_n^{(0)(k)^{-1}} \underline{\epsilon}_n^{(0)(k)} + \ln(|\mathbf{E}_n^{(1)(k)}|/|\mathbf{E}_n^{(0)(k)}|)$

4. $h_n = h_{n-1} + \Delta h_n, \begin{array}{l} < -2\ln A \rightarrow H_1 \\ > -2\ln B \rightarrow H_0 \end{array}$

The performance of the classifier can be evaluated with two key measures of performance: the number of correct classifications and the length of data (length of time) required to make a classification. By varying different properties of the classifier and environment (e.g., signal-to-noise ratio), the effects of these different properties on the classifier can be evaluated using the two performance measures. The remainder of this chapter describes the simulations and results for varying conditions of the classifier and environment.

E. SIMULATION SETUP AND PARAMETERS

In order to test the proof of concept for the multirate sequential classifier multiple simulations were conducted. The test data consisted of recordings of sound from one propellor plane and three different A-10 jet aircraft, with noise added. The parameters that were varied during the simulations were the signal-to-noise ratio (SNR) associated with the training and target data sequences, and the length of the training sequences used to develop the classifier parameters. Simulations were conducted to compare the single-channel, multichannel and multirate cases.

1. Test Data Characteristics

This subsection describes the characteristics of the test data. The original audio data sequences were recorded at 44,100 Hz. The data was initially low-pass filtered and downsampled by a factor of 10 to achieve a sampling rate of 4,410 Hz. (The observed data for the simulation is then further decimated.) A spectral estimate was then computed for each of the four data sequences prior to adding any additional noise, in order to determine any unique characteristics. Figure 5.2 shows the spectral plots of each sequence and Table 5.3 summarizes the dominant frequencies, and a secondary frequency if noticeable, that were measured from the spectral plots.

From Fig. 5.2, it can be seen that the propellor plane has secondary frequency at a lower frequency than the peak frequency of 192.0 Hz. A secondary frequency is

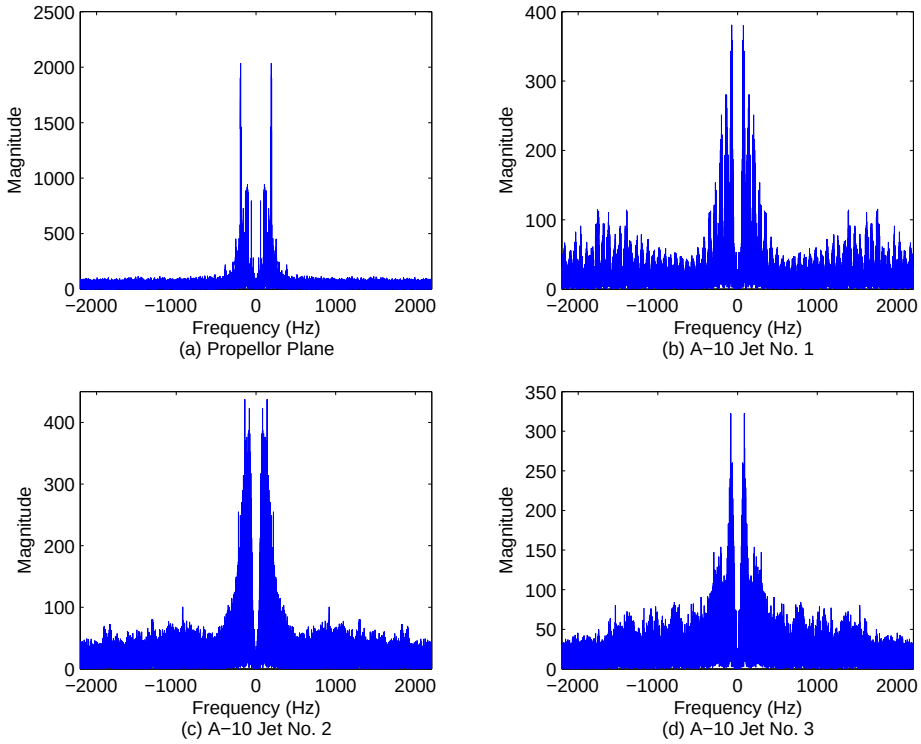


Figure 5.2. Spectral Plots

Table 5.3. Spectral Information for Data Sequences

	Propellor Plane	A-10 Jet No. 1	A-10 Jet No. 2	A-10 Jet No. 3
Dominant Freq (Hz)	192.0	145.0	82.5	84.6
Secondary Freq (Hz)	106.9	N/A	N/A	210.4

also seen on the A-10 Jet No. 3 at 210.4 Hz. The A-10 Jet No. 1 has many peaks close to each other. However, the magnitude of these peaks diminish as the distance from the main peak increases. Thus they have not been considered as secondary peaks.

In addition to the spectral estimate of the test data, the autocorrelation of each data sequence was calculated to examine any significant differences in correlation. The autocorrelation plots were used to assist in selecting appropriate filter orders for calculating the prediction orders used in the classifier. One can see from Fig. 5.3 that the three A-10 Jet planes have similar, but not identical, correlation structure within the first ten samples (dashed lines in Fig. 5.3.) This similar structure led to choosing a prediction filter order larger than ten, as discussed in the next subsection.

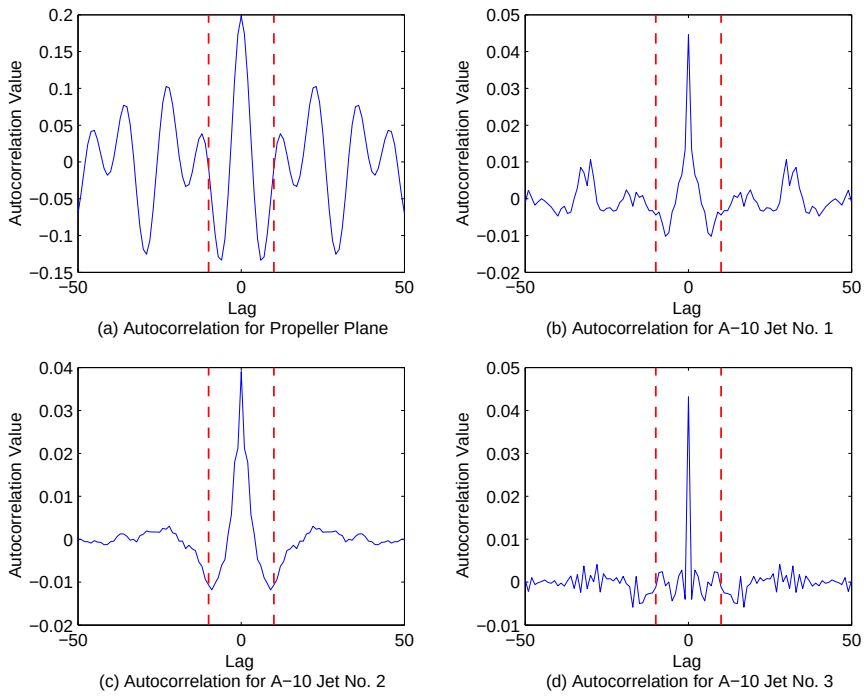


Figure 5.3. Estimated Autocorrelation Function for Aircraft Data at 4,410 Hz sampling rate. Dashed lines show ±10 lag values.

2. Simulation Models

For the simulations conducted, three scenarios were considered. These scenarios considered classification using a single-channel system, the multichannel system and the multirate system. All scenarios began with the recorded data sampled at 4410 Hz. For the single-channel scenario, the recorded data was decimated by a factor of two, resulting in a sampling rate of 2205 Hz. This was done to ensure that the highest rate in all three scenarios would be the same. Additive white Gaussian noise with a mean of zero and a variance of σ_η^2 was then added to achieve a desired Signal-to-Noise ratio set by the simulation parameters. The proper selection of the value of σ_η^2 is described below. The single-channel scenario is shown in Fig. 5.4.

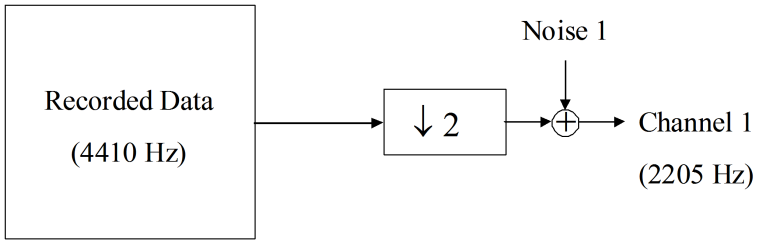


Figure 5.4. Observation Model for the Single Channel Scenario.

The simulation model for the multichannel scenario is similar to the single-channel scenario, with the exception that after the recorded data is decimated by a factor of two, the signal is split into two channels. Noise was then added to each channel to achieve a desired SNR. The model for the multichannel scenario is shown in Fig. 5.5.

For the multirate scenario, Fig. 5.6, the high rate channel was decimated by a factor of 2 and the low rate channel was decimated by a factor of 3. Noise was then added to each channel to achieve a desired signal-to-noise ratio.

In order to produce the desired SNR for these models, the noise power must be calculated using the desired SNR in decibels and the estimated power of the appropriate channel. If the output of the decimator is $s[m_x]$, where m_x is the time

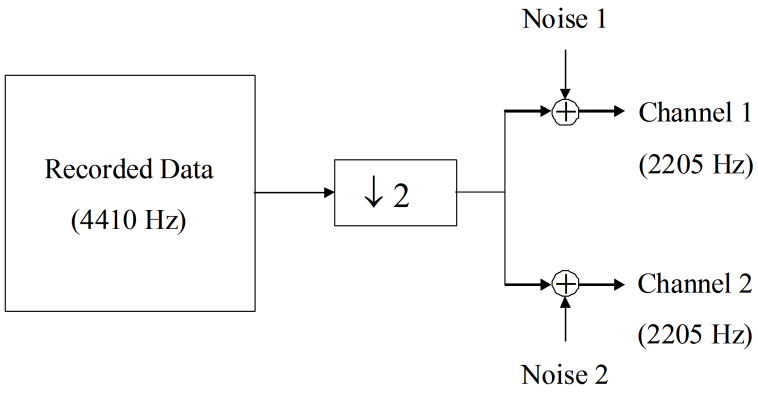


Figure 5.5. Observation Model for the Multichannel Scenario.

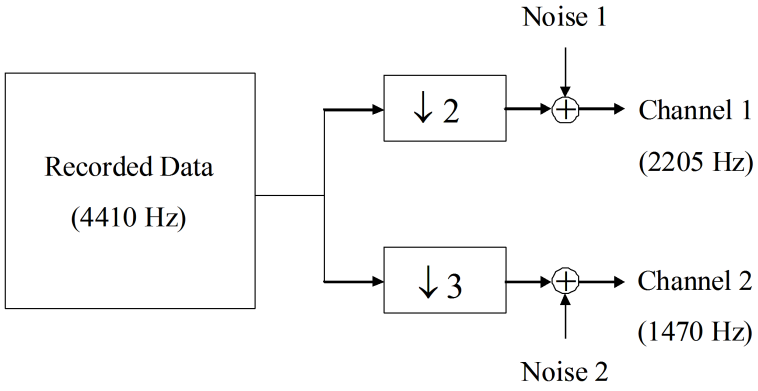


Figure 5.6. Observation Model for the Multirate Scenario.

index for the decimated signal, $\eta[m_x]$ is the noise signal, and the observed signal $x[m_x]$ is

$$x[m_x] = s[m_x] + \eta[m_x],$$

then the noise power, σ_η^2, can be calculated as

$$\sigma_\eta^2 = \frac{\frac{1}{m_x}\sum_{i=0}^{m_x-1} s^*[i]s[i]}{10^{-\mathrm{SNR}/10}}.$$

Once the noise power has been calculated, this value can be used to generate the appropriate noise signal, $\eta[m_x]$.

3. Prediction Coefficients

When calculating the prediction coefficients for the training data, an appropriate filter order has to be chosen. In general, the filter order should be large enough to exploit differences in correlation between the two classes of signals being tested. It should be small enough, however, so that it does not add excessive computational burden to the simulation and does not become sensitive to small artifacts of the training data. In addition, as the filter order increases, the determinant of the prediction error covariance, which measures the quality of the prediction, typically approaches a value such that any further increases in filter order do not result in an appreciable improvement in performance.

From Fig. 5.3, it can be seen that the propellor plane is uniquely different in its correlation structure, but the three A-10 jets have similar structures for lags smaller then ten. Thus choosing an order for the linear prediction filter less than ten would probably not provide enough uniqueness in the three A-10 jet parameters to be able to adequately classify the different planes. However, the correlation shapes at lags larger than ten are sufficiently dissimilar to exploit these differences in correlations. Although larger filter orders could have been used, the filter order was chosen to be 40 to minimize excess computations.

The prediction coefficients and prediction error variances were calculated in the manner described in Chapter IV, using the multichannel Levinson recursion. Since the decimation rates for the two channels in the multirate scenario are two and three, the system period, K, is equal to 6. In addition, there are a total of five observations that occur within one system period, three for the high-rate channel and two for the low-rate channel. The sampling pattern is shown in Fig. 5.7. It is assumed the data is aligned so that at the initial observation, both channels produce observations.

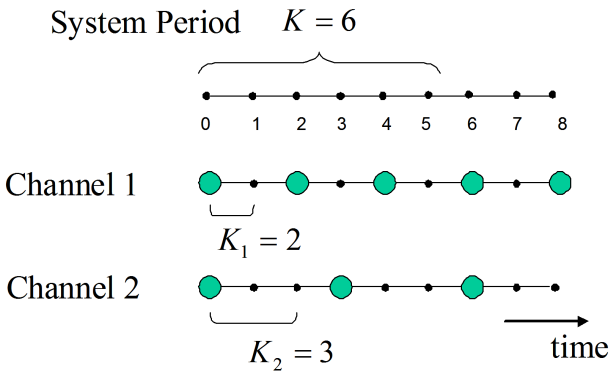


Figure 5.7. Illustration of Decimation for the Multirate Scenario.

4. Training Data

When calculating the prediction coefficients, a set of training data for each class had to be used. This data consisted of a portion of the total data signal used for each class, as shown shaded in Fig. 5.8. The length of this portion was varied according to the scenario parameters. The rest of the signal was then used for the testing portion of the simulation.

During the testing phase, the simulation starts at the beginning of the test portion of the input signal and continues to take observations until a classification is made. This is considered to be one "trial." A second trial is then begun starting with the observation corresponding to the data point just after the data point at which the first classification was made, as shown in Fig. 5.8. This process is continued

until 50 trials are conducted. The length of the data was sufficient to ensure that the simulation was able to conduct 50 trials before reaching the end of the data.

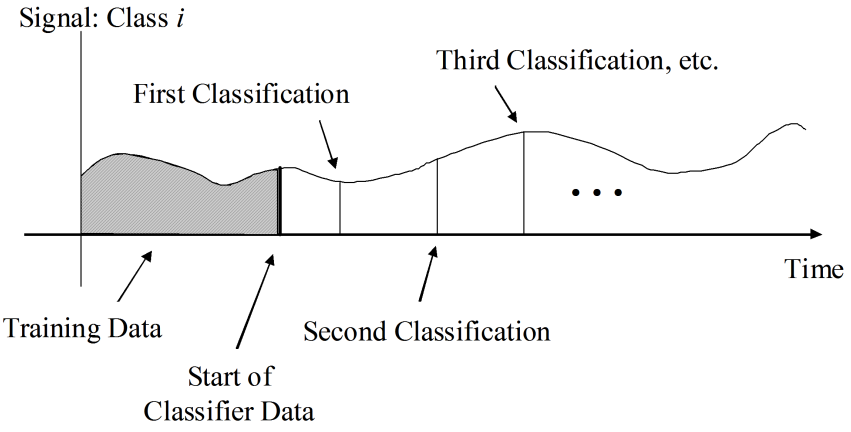


Figure 5.8. Training Data.

Although the sequential classifier can in theory classify data based on as little as one single observation, in practice it is better to allow some minimum number of observations to be collected before a classification is attempted. In these experiments, this minimum data length was linked to the length of the linear prediction filter. In particular, the algorithm was designed not to classify the target until the number of observations exceeded the filter order by one observation. It was decided to implement the algorithm this way to avoid any errant classifications that might occur while the algorithm collects the observations needed to make the predictions at full order.

F. SIMULATION RESULTS AND ANALYSIS

The results for the multirate scenarios are presented in the following subsection; the key parameters are analyzed in the last subsection.

1. Multirate Scenario Results

Since there were four classes of data (one propellor and three jet planes) the classifier was tested using two classes at a time. This resulted in six sets of experiments, whose results are listed in Tables 5.4 through 5.9 of this section.

In order to make comparisons, specific parameters were varied for the simulations. These were the signal-to-noise ratio (SNR) of the target data and the length of data sequences used to calculate class parameters. In all cases, the training data was assumed to be noiseless but white noise was added to the test data to achieve a desired SNR.

Table 5.4. Classification Results: Propellor vs. A-10 Jet No. 1

SNR	Data Training Length	Propellor No. of Correct Classifications (out of 50)	Jet No. 1 No. of Correct Classifications (out of 50)	Average Time to Classify Propellor (msec)	Average Time to Classify Jet No. 1 (msec)
40 dB	6250	50	50	9.52	9.52
	2500	50	50	9.52	9.52
20 dB	6250	50	50	9.52	9.59
	2500	50	39	9.52	9.52
10 dB	6250	50	21	9.52	9.52
	2500	50	10	9.52	9.52

Table 5.4 shows the classification results for the propellor plane versus the A-10 Jet No. 1. This table lists both the number of correct classifications for each target as well as the average time to classify the target. The average time of 9.52 msec listed in the table corresponds to the minimum observation condition of 41 samples (see Subsection V.E.4). It can be seen from Table 5.4 that the classifier makes its decision almost immediately after reaching the full filter order under all cases. Thus,

for the case of the propellor plane versus Jet No. 1, the advantage of a sequential classifier is not realized. This is not true for all the test scenarios.

From Table 5.4, one can see that the algorithm is able to classify the propellor under all conditions, and almost immediately after collecting enough data to reach full order. In an environment with an SNR of 40 dB, the classifier is able to classify the propellor plane nearly perfectly. As the SNR is reduced to 20 dB, the success rate reduces to a little below 80% for the short training data length and under 15% for the high training data length. With the middle length training data classification is perfect. Proportionately similar reductions occur as the SNR is dropped further to 10 dB. The results degrade significantly at an SNR of 10 dB. Namely, the classifier tends to classify both signals as a propellor plane; this puts into question whether the classifier appropriately classifies the propellor plane at 10 dB or defaults to the propellor plane because of the signal degradation. In almost all cases classification occurred with the minimum allowable number of observations. For SNR below 5 dB, classification degraded severely. Therefore, the results are not presented in these tables.

Table 5.5. Classification Results: Propellor vs. A-10 Jet No. 2

SNR	Data Training Length	Propellor No. of Correct Classifications (out of 50)	A-10 No. 2 No. of Correct Classifications (out of 50)	Average Time to Classify Propellor (msec)	Average Time to Classify A-10 No. 2 (msec)
40 dB	6250	50	50	9.56	9.52
	2500	50	50	9.53	9.52
20 dB	6250	50	36	9.52	9.53
	2500	50	8	9.52	9.53
10 dB	6250	50	2	9.52	9.52
	2500	50	0	9.52	9.52

The results of classification for the propellor plane and the A-10 Jet No. 2 are contained in Table 5.5. Similar to the testing between the propellor plane and

the A-10 Jet No. 1, the algorithm was able to correctly classify the propellor plane nearly perfectly under all conditions simulated. In addition, the classification occurred almost immediately after reaching the minimum number of observations allowed. The number of correct classifications for Jet No. 2 was slightly less than that for Jet No. 1. As with Jet No. 1, the performance of Jet No. 2 dropped appreciably as the SNR was reduced to 20 dB. Under the conditions with an SNR of 10 dB, the classifier performed poorly when trying to classify Jet No. 2.

Table 5.6. Classification Results: Propellor vs. A-10 Jet No. 3

SNR	Data Training Length	Propellor No. of Correct Classifications (out of 50)	A-10 No. 3 No. of Correct Classifications (out of 50)	Average Time to Classify Propellor (msec)	Average Time to Classify A-10 No. 3 (msec)
40 dB	6250	50	50	9.52	9.52
	2500	50	50	9.52	9.52
20 dB	6250	50	45	9.52	9.53
	2500	50	35	9.52	9.53
10 dB	6250	50	2	9.52	9.52
	2500	50	0	9.52	9.52

For the experiments conducted between the propellor plane and the A-10 Jet No. 3 shown in Table 5.6, the algorithm once again was able to classify the propellor plane one hundred percent of the time and with the minimum number of observations allowed. The results for Jet No. 3 were somewhat between the result for the other two A-10 jets. For the simulations with an SNR of 40 dB, the algorithm was able to classify Jet No. 3 perfectly. As the SNR was reduced to 20 dB, the performance dropped to between 70% and 90%. The performance then significantly dropped to almost no correct classifications with an SNR of 10 dB.

Besides comparing the jet planes to the propellor plane, it was also of interest to see if the classifier could discriminate between the different types of jet planes.

Table 5.7. Classification Results: A-10 Jet No. 1 vs. A-10 Jet No. 2

SNR	Data Training Length	A-10 No. 1 No. of Correct Classifications (out of 50)	A-10 No. 2 No. of Correct Classifications (out of 50)	Average Time to Classify A-10 No. 1 (msec)	Average Time to Classify A-10 No. 2 (msec)
40 dB	6250	49	47	10.01	9.94
	2500	11	50	9.68	9.52
20 dB	6250	43	23	9.68	9.67
	2500	19	42	9.65	9.54
10 dB	6250	43	16	9.54	9.57
	2500	15	37	9.52	9.52

This is obviously a more difficult problem than discriminating between a jet plane and a propellor plane.

When comparing the different jets, the results were more balanced than those seen comparing the jets with the propellor plane. For the experiments conducted comparing Jet No. 1 and Jet No. 2, shown in Table 5.7, percentages of correct classification greater than eighty percent were typical for about half of the scenarios. The rest of the results ranged from 0% to about 50%. The time needed for the algorithm to make a classification decision is somewhat higher than the time needed when comparing a jet plane to the propellor plane.

Table 5.8 contains the results for the experiments conducted comparing the A-10 Jet No. 1 with Jet No. 3. From this table one can see that the algorithm was able to successfully classify Jet No. 1 correctly for the environments with SNRs of 20 dB and 10 dB, while Jet No. 3 performed poorly under these conditions. The results were opposite for the case of an SNR of 40 dB. Classifications in the 20 dB and 10 dB cases were made almost immediately after reaching the minimum allowable length. For the 40 dB case the classification time was significantly longer.

Results for comparing the A-10 Jet No. 2 and the A-10 Jet No. 3 are contained in Table 5.9. From this table one can see that nearly perfect results for classifying

Table 5.8. Classification Results: A-10 Jet No. 1 vs. A-10 Jet No. 3

SNR	Data Training Length	A-10 No. 1 No. of Correct Classifications (out of 50)	A-10 No. 3 No. of Correct Classifications (out of 50)	Average Time to Classify A-10 No. 1 (msec)	Average Time to Classify A-10 No. 3 (msec)
40 dB	6250	11	50	10.56	9.52
	2500	5	50	9.53	9.52
20 dB	6250	50	1	9.52	9.66
	2500	50	2	9.52	9.52
10 dB	6250	50	0	9.52	9.52
	2500	50	0	9.52	9.52

Table 5.9. Classification Results: A-10 Jet No. 2 vs. A-10 Jet No. 3

SNR	Data Training Length	A-10 No. 2 No. of Correct Classifications (out of 50)	A-10 No. 3 No. of Correct Classifications (out of 50)	Average Time to Classify A-10 No. 2 (msec)	Average Time to Classify A-10 No. 3 (msec)
40 dB	6250	40	49	12.48	11.40
	2500	50	49	13.39	11.65
20 dB	6250	49	4	9.53	9.56
	2500	50	0	9.52	9.52
10 dB	6250	50	0	9.52	9.52
	2500	50	0	9.52	9.52

Jet No. 2 were obtained. Classifications occurred almost immediately after reaching full order for the 20 dB and 10 dB cases. While more time was needed for the 40 dB cases. For Jet No. 3 nearly perfect classification results were obtained for the 40 dB cases. This performance fell to nearly 0% for the 20 dB ad 10 dB cases. Similar to the experiments with the propellor plane and Jet No. 1, as the SNR is dropped significantly, degradation only in Jet No. 3 tends to show that the classifier is skewed towards Jet No. 2 questioning whether the results for Jet No. 2 at SNRs of 20 dB and 10 dB are reliable.

In another set of experiments, the classifier designed using the propellor and Jet No. 1 was tested on data from all four classes (propellor, Jet No. 1, Jet No. 2 and Jet No. 3). When tested on Jet No. 2 and Jet No. 3, the classification was considered to be correct if the classifier chose the "Jet" category, and incorrect if the classifier classified these jet planes as propellor planes.

Results of these experiments are listed in Table 5.10. From these results, it can be seen that the classifier was able to correctly classify the A-10 Jets No. 2 and No. 3 as jets perfectly for the 40 dB scenarios. When the SNR was reduced to 20 dB, the A-10 Jet No. 3 was classified nearly perfectly, while the A-10 Jet No. 2 had at most a classification rate of about 60%. Under the 10 dB conditions, the classifier performed poorly when trying to classify the A-10 Jets No. 2 and No. 3.

2. Analysis of Single-channel vs. Multirate vs. Multichannel Simulations

Experiments identical to those conducted for the multirate data were conducted for the single channel and multichannel scenarios. In most situations, it was possible to adequately classify the target in these other scenarios. Therefore, to compare the effectiveness of the methods, the average number of data samples needed to make a classification were also compared. Tables 5.11 through 5.13 show results for some of the experiments for an environment with an SNR of 40 dB. Table 5.11 shows that under these conditions, the classifier was able to correctly classify the propellor

Table 5.10. Classification Results: Class 0 = Propellor and Class 1 = A-10 Jet No. 1

SNR	Data Training Length	Propellor No. of Correct Classifications (out of 50)	Jet No. 1 No. of Correct Classifications (out of 50)	Jet No. 2 No. of Correct Classifications (out of 50)	Jet No. 3 No. of Correct Classifications (out of 50)	Average Time to Classify Propellor (msec)	Average Time to Classify Jet No. 1 (msec)	Average Time to Classify Jet No. 2 (msec)	Average Time to Classify Jet No. 3 (msec)
40 dB	6250	50	50	50	50	9.52	9.52	9.52	9.52
	2500	50	50	30	50	9.52	9.52	10.24	9.52
20 dB	6250	50	50	32	47	9.52	9.52	9.57	9.53
	2500	50	39	3	41	9.52	9.75	9.55	9.54
10 dB	6250	50	21	2	9	9.52	9.52	9.52	9.52
	2500	50	10	0	10	9.52	9.52	9.52	9.53

and Jet No. 2 in all these scenarios. In addition, it can be seen that the time needed to make the classification is roughly the same for the single-channel, multichannel and multirate cases.

Table 5.11. Classification Results for Propellor and Jet No. 2

Scenario	Data Training Length	Propellor No. of Correct Classifications (out of 50)	Jet No. 2 No. of Correct Classifications (out of 50)	Average Time to Classify Propellor (msec)	Average Time to Classify Jet No. 2 (msec)
Single-Channel	6250	50	50	10.12	9.56
	2500	50	50	9.56	9.52
Multichannel	6250	50	50	9.52	9.52
	2500	50	50	9.76	10.28
Multirate	6250	50	50	9.53	9.52
	2500	50	50	9.52	9.53

In table 5.12, it can be seen that ability to distinguish between Jet No. 1 and Jet No. 2 becomes more difficult than between the propellor plane and jets for all three scenarios. The multirate classifier, however, is able to make its classifications in a shorter amount of time.

In the example of the A-10 Jet No. 2 versus the A-10 Jet No. 3, Table 5.13, the classifiers again performed fairly well. Although, the amount of time necessary to make classification decisions for all three cases was slightly higher than for some of the other simulations, the multirate classifier on average needed less time to make its classification.

From these tables it can be seen that the multirate case has some advantage over the single-channel case. In low noise environments where the single-channel, multirate and multichannel scenarios all had a high success rate of proper classification, the multirate scenario was able to classify the proper signal in a shorter amount of time. It was noticed that the multichannel channel case at times needed more time

Table 5.12. Classification Results for Jet No. 1 and Jet No. 2

Scenario	Data Training Length	Jet No. 1 No. of Correct Classifications (out of 50)	Jet No. 2 No. of Correct Classifications (out of 50)	Average Time to Classify Jet No. 1 (msec)	Average Time to Classify Jet No. 2 (msec)
Single-Channel	6250	50	36	9.90	35.78
	2500	45	49	15.96	22.78
Multichannel	6250	50	35	9.89	32.42
	2500	46	50	15.37	22.51
Multirate	6250	49	47	10.02	9.94
	2500	11	50	9.68	9.52

Table 5.13. Classification Results for Jet No. 2 and Jet No. 3

Scenario	Data Training Length	Jet No. 2 No. of Correct Classifications (out of 50)	Jet No. 3 No. of Correct Classifications (out of 50)	Average Time to Classify Jet No. 2 (msec)	Average Time to Classify Jet No. 3 (msec)
Single-Channel	6250	35	50	20.10	10.74
	2500	50	49	15.13	13.08
Multichannel	6250	37	50	17.29	11.31
	2500	50	48	12.05	15.18
Multirate	6250	40	49	12.48	11.40
	2500	50	49	13.48	11.65

to make a classification decision than the multirate case. Although there is more data available in the multichannel scenario, the offset observations of the multirate scenario adds more statistical information to the solution.

3. Key Parameters for the Multirate Case

There were three main characteristics of the data that were analyzed for their effect on the classification performance. They were the dominant frequency of the target data, the signal-to-noise ratio and the training data length. In addition, the performance between single-channel, multirate and multichannel observations were compared. Each is described individually in the following subsubsections, however, their impact is not independent of each other. So, if noticeable, their relative impact on each other is discussed.

a. Dominant Frequency

While analyzing the data, it became apparent that the classifier performed much better when one target data class had a dominant frequency higher than that of the other class to which it was being compared. This was especially noticeable in the high noise environment scenarios. Under the high noise conditions, the jets were not classified correctly with any degree of success against the propellor plane. Only under the low noise condition ($SNR = 10$ dB) did the jets display moderate success with classification against the propellor plane. Likewise, Jet No. 1 had a higher correct classification rate when attempting to classify against Jets No. 2 and No. 3. And, Jet No. 2 performed better against Jet No. 3.

b. Signal-to-Noise Ratio

By analyzing the simulation data, it became noticeable that under low noise conditions ($SNR = 40$ dB), the three A-10 jets were classified properly against the propellor plane. As the training sample length was reduced, the performance degraded dramatically. In addition, the higher noise conditions reduced the

performance of the jets significantly when compared to the propellor plane. It appears that the dominant frequency has a large effect on the outcome of the classifier, in any condition other than ideal.

c. Training Data Length

As might be anticipated, the longer the training data length, the better the classifier typically performed. Not only did the classification rate generally increase, but the number of samples needed to make a proper classification was reduced. For the audio data used, however, when the size of the training data was increased to 12500 samples the performance of the classifier actually degraded. It was determined that at this length the data available for testing became statistically different enough from the training data and caused a degradation in the performance of the classifiers.

G. CONCLUSION

In this chapter, it was shown that with some modifications to the single-channel classifier, one can implement a sequential classifier that accepts multirate data. In addition, using the multirate Levinson recursion developed in Chapter IV; one can derive the needed class parameters for the multirate sequential classifier. The multirate sequential classifier can handle multiple channels of observations occurring at different sampling intervals.

Simulations were conducted to analyze the effectiveness of the multirate sequential classifier compared with a single-channel classifier and a multichannel classifier where all channels were sampled at the same rate. It was seen that under low noise environments all three classifiers performed well, but the multirate classifier was able to make classification decisions in a shorter amount of time due to the added information from the offset observations. The amount of addiitonal information contained in the multirate classifier is dependent upon the disparity between channel sampling rates.

THIS PAGE INTENTIONALLY LEFT BLANK

VI. CONCLUSIONS AND FUTURE WORK

The goal of this work was to further develop the theory and applications of statistical multirate signal processing. A summary of the work and contributions of this thesis is provided here. This is followed by a section of suggestions for further research.

A. CONCLUSIONS

In Chapter II, the foundation necessary to describe multirate systems is established. Part of this foundation requires defining some key terms applicable to multirate systems, such as fundamental rate, system period and system phase. These terms are used to explicitly describe the periodic nature of the multirate system and allow for the appropriate selection of values for the periodic components in the multirate system. The definitions for describing the statistical characterizations of multirate signals and building blocks (e.g., decimator and expander) are provided. Compact matrix forms of the decimator, expander and filters, which take advantage of Kronecker product relations, are presented. Finally, the LPTV filter bank of [Ref. 17] is generalized to apply to systems with different input and output rates.

In Chapter III, the explicit direct form of the multirate optimal estimator is developed. This optimal estimator uses multiple-input signals observed at different sampling rates to estimate a desired signal. In addition, a recursive innovations form of this multirate optimal estimator is derived. This innovations form of the optimal estimator separates the direct form optimal filters into modified optimal filters and cross filters. These cross filters remove any information from one signal that is contained in the other signals, in an ordered fashion. In essence, a new set of input signals that are mutually orthogonal are derived and used as the inputs to the modified optimal filters. Using the recursive form of the optimal filter allows

one to calculate the relative change in performance in the optimal estimator when input signals are added or removed from the system. Through simulations, it is demonstrated that optimal filtering of multiple multirate channels for an AR process and a sinusoidal process can provide improved performance over optimal filtering using a single channel, even if the secondary channel when used by itself has high error variances.

In Chapter IV, the optimal filtering problem is specialized for the case of optimal multirate linear prediction. The multirate Normal equations are derived for a multirate system with multiple input and output signals, which are observed at different sampling rates. For a multirate system with a system period K, there are up to K distinct sets of prediction coefficients and error covariance matrices that apply in a periodic fashion. An efficient method for calculating the multirate linear prediction coefficients and error variances is developed through the use of the multichannel Levinson recursion and generalized triangular UL factorization. A specific algorithm for the generalized triangular factorization is included in Appendix E.

In Chapter V, a multirate sequential classifier is derived starting from the basic theory of sequential hypothesis testing. It is shown that classifier parameters needed for implementing the multirate sequential classifier are the same as those for multirate linear prediction. A multirate sequential classifier is then implemented and tested using audio files of a propellor plane and three A-10 jet aircraft. The experiments tested the classifier performance in selecting between the propellor plane and jet aircraft as certain system parameters are changed. These parameters consisted of the signal-to-noise ratios of the observed signal and the length of the training data. In addition, the performance of the multirate classifier is compared to that of the

single-channel and multichannel classifiers using similar data. These experiments showed that improved performance can be obtained by using two channels of observation data even if one channel is sampled at a lower rate.

B. TOPICS FOR FURTHER INVESTIGATION

Multirate signal processing continues to be an active area of research and many opportunities exist for further research, particularly in the area of *statistical* signal processing. The optimal multirate filter developed in this dissertation is an FIR filter. The infinite impulse response (IIR) filter was not considered here; however, multirate theory applies equally well to the IIR case and also would provide significant contributions to the theory of multirate signal processing.

Another opportunity for further work on multirate systems is in the area of lattice structures. Some work on lattice structures has already been researched, but extensive development of lattice theory as it applies to multirate linear prediction can lead to more convenient representation and implementation. In addition, this dissertation focused on optimal solutions using time-domain characteristics. Solutions to optimal filtering of multirate systems using frequency-domain techniques would further expand the theory and utility of multirate statistical signal processing. Research into filter design of periodically correlated scalar time series using a spectral factorization algorithm for wide-sense stationary vector processes has already been started by Spurbeck and Scharf [Ref. 29].

Research from a companion thesis [Ref. 50] has extended the concepts of optimal multirate filtering to two-dimensional (2-D) signal processing and high-resolution image reconstruction. The reader of this thesis may want to refer to [Ref. 50] for some additional topics not treated here.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A. POWER SPECTRAL DENSITY FOR MULTIRATE PROCESSES

This appendix contains the derivations of the power spectral density functions for the the decimator, expander and filters presented in Chapter II. The standard equation for the power spectral density function of a stationary process

$$S(\omega) = \sum_{l=-\infty}^{\infty} R[l]e^{-j\omega l} \tag{A.1}$$

can be applied to the input and output correlations independently. However, any difference between sampling rates of the input and output signals makes it impossible to develop a power spectral density function for the cross-correlation dependent on a lag only. Therefore it is necessary to use a two-dimensional power spectral density function. This function is defined as

$$S^{2D}(\lambda_1, \lambda_0) = \sum_{n_1=-\infty}^{\infty} \sum_{n_0=-\infty}^{\infty} R[n_1, n_0]e^{-j\lambda_1 n_1}e^{-j\lambda_0 n_0}. \tag{A.2}$$

For the input to a decimator, expander or filter, the power spectral density is defined as

$$S_x^{2D}(\lambda_1, \lambda_0) = \sum_{n_1=-\infty}^{\infty} \sum_{n_0=-\infty}^{\infty} R_x[n_1, n_0]e^{-j\lambda_1 n_1}e^{-j\lambda_0 n_0}. \tag{A.3}$$

Using the time-lag form (when appropriate), the power spectral density function becomes

$$S_x^{2D}(\lambda, \omega) = \sum_{n=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} R_x[n; l]e^{-j\lambda n}e^{-j\omega l}, \tag{A.4}$$

which reduces to (A.1) when the signal is wide sense stationary.

A. DECIMATION

The operation of decimation is depicted in Fig. A.1 and defined as

$$y[m_y] = x[Lm_y] \quad \text{for } m = \ldots, -1, 0, 1, \ldots.$$

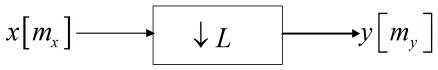


Figure A.1. Decimator

The correlation functions for the decimator are as shown in Table 2.2 of Chapter II. The output power spectral density function for decimation is

$$\begin{aligned} S_y^{2D}(\lambda_1, \lambda_0) &= \sum_{m_y=-\infty}^{\infty} \sum_{m'_y=-\infty}^{\infty} R_y[m_y, m'_y] e^{-j\lambda_1 m_y} e^{-j\lambda_0 m'_y} \\ &= \sum_{m_y=-\infty}^{\infty} \sum_{m'_y=-\infty}^{\infty} R_x[Lm_y, Lm'_y] e^{-j\lambda_1 m'_y} e^{-j\lambda_0 m_y} \\ &= \sum_{m_y=-\infty}^{\infty} \sum_{m'_y=-\infty}^{\infty} R_x[Lm_y, Lm'_y] e^{-j(\lambda_1/L) Lm'_y} e^{-j(\lambda_0/L) Lm_y} \\ S_y^{2D}(\lambda_1, \lambda_0) &= S_x^{2D}\left(\frac{\lambda_1}{L}, \frac{\lambda_0}{L}\right) \end{aligned}$$

In the time-lag representation, the output power spectral density function is defined

as

$$\begin{aligned}
S_y^{2D}(\lambda,\omega) &= \sum_{m_y=-\infty}^{\infty}\sum_{l_y=-\infty}^{\infty} R_y[m_y;l_y]e^{-j\lambda m_y}e^{-j\omega l_y} \\
&= \sum_{m_y=-\infty}^{\infty}\sum_{l_y=-\infty}^{\infty} R_x[Lm_y;Ll_y]e^{-j\lambda m_y}e^{-j\omega l_y} \\
&= \sum_{m_y=-\infty}^{\infty}\sum_{l_y=-\infty}^{\infty} R_x[Lm_y;Ll_y]e^{-j(\lambda/L)Lm_y}e^{-j(\omega/L)Ll_y} \\
S_y^{2D}(\lambda,\omega) &= S_x^{2D}\left(\frac{\lambda}{L},\frac{\omega}{L}\right)
\end{aligned}$$

or in lag representation only

$$\begin{aligned}
S_y(\omega) &= \sum_{l_y=-\infty}^{\infty} R_y[l_y]e^{-j\omega l_y} \\
&= \sum_{l_y=-\infty}^{\infty} R_x[Ll_y]e^{-j\omega l_y} \\
&= \sum_{l_y=-\infty}^{\infty} R_x[Ll_y]e^{-j(\omega/L)Ll_y} \\
S_y(\omega) &= S_x\left(\frac{\omega}{L}\right)
\end{aligned}$$

The cross-power spectral density function is given by

$$\begin{aligned}
S_{xy}^{2D}(\lambda_x,\lambda_y) &= \sum_{m_x=-\infty}^{\infty}\sum_{m_y=-\infty}^{\infty} R_{xy}[m_x,m_y]e^{-j\lambda_x m_x}e^{-j\lambda_y m_y} \\
&= \sum_{m_x=-\infty}^{\infty}\sum_{m_y=-\infty}^{\infty} R_x[m_x,Lm_y]e^{-j\lambda_x m_x}e^{-j\lambda_y m_y} \\
&= \sum_{m_x=-\infty}^{\infty}\sum_{m_y=-\infty}^{\infty} R_x[m_x,Lm_y]e^{-j\lambda_x m_x}e^{-j(\lambda_y/L)Lm_y} \\
S_{xy}^{2D}(\lambda_x,\lambda_y) &= S_x^{2D}\left(\lambda_x,\frac{\lambda_y}{L}\right)
\end{aligned}$$

In the time-lag representation, the cross-power spectral density function is defined as

$$
\begin{aligned}
S_{xy}^{2D}(\lambda_x, \omega_y) &= \sum_{m_x=-\infty}^{\infty} \sum_{l_y=-\infty}^{\infty} R_{xy}[m_x; l_y] e^{-j\lambda_x m_x} e^{-j\omega_y l_y} \\
&= \sum_{m_x=-\infty}^{\infty} \sum_{l_y=-\infty}^{\infty} R_x[m_x; Ll_y - (L-1)m_x] e^{-j\lambda_x m_x} e^{-j\omega_y l_y} \\
&= \sum_{m_y=-\infty}^{\infty} \sum_{l_y=-\infty}^{\infty} R_x[m_x; Ll_y - (L-1)m_x] e^{-j\lambda_x m_x} e^{-j\left(\frac{L-1}{L}\omega_y\right)m_x} e^{-j(\omega_y/L)Ll_y} e^{-j\left(-\frac{L-1}{L}\omega_y\right)m_x} \\
&= \sum_{m_y=-\infty}^{\infty} \sum_{l_y=-\infty}^{\infty} R_x[m_x; Ll_y - (L-1)m_x] e^{-j\left(\lambda_x + \frac{L-1}{L}\omega_y\right)m_x} e^{-j(\omega_y/L)(Ll_y - (L-1)m_x)} \\
S_{xy}^{2D}(\lambda_x, \omega_y) &= S_x^{2D}\left(\lambda_x + \frac{L-1}{L}\omega_y, \frac{\omega_y}{L}\right)
\end{aligned}
$$

B. EXPANSION

The operation of expansion is depicted in Fig. A.2 and defined as

$$
y[m] = \begin{cases} x[m/I] & m \text{ div } I \\ 0 & \text{otherwise} \end{cases}.
$$

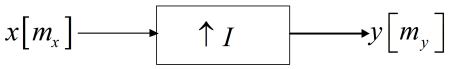


Figure A.2. Expander

The correlation functions for the expander are as shown in Table 2.3 in Chapter

II. The output power spectral density function for expansion is

$$\begin{aligned}
S_y^{2D}(\lambda_1,\lambda_0) &= \sum_{m_y=-\infty}^{\infty}\sum_{m_y'=-\infty}^{\infty} R_y[m_y,m_y']e^{-j\lambda_1 m_y}e^{-j\lambda_0 m_y'} \\
&= \sum_{m_y=-\infty}^{\infty}\sum_{m_y'=-\infty}^{\infty} R_x[m_y/I,m_y'/I]e^{-j\lambda_1 m_y}e^{-j\lambda_0 m_y'} \\
&= \sum_{m_y=-\infty}^{\infty}\sum_{m_y'=-\infty}^{\infty} R_x[m_y/I,m_y'/I]e^{-j(I\lambda_1) m_y/I}e^{-j(I\lambda_0) m_y'/I} \\
S_y^{2D}(\lambda_1,\lambda_0) &= S_x^{2D}\left(I\lambda_1, I\lambda_0\right)
\end{aligned}$$

In the time-lag representation, the output power spectral density function is defined as

$$\begin{aligned}
S_y^{2D}(\lambda,\omega) &= \sum_{m_y=-\infty}^{\infty}\sum_{l_y=-\infty}^{\infty} R_y[m_y;l_y]e^{-j\lambda m_y}e^{-j\omega l_y} \\
&= \sum_{m_y=-\infty}^{\infty}\sum_{l_y=-\infty}^{\infty} R_x[m_y/I;l_y/I]e^{-j\lambda m_y}e^{-j\omega l_y} \\
&= \sum_{m_y=-\infty}^{\infty}\sum_{l_y=-\infty}^{\infty} R_x[m_y/I;l_y/I]e^{-j(I\lambda) m_y/I}e^{-j(I\omega) l_y/I} \\
S_y^{2D}(\lambda,\omega) &= S_x^{2D}\left(I\lambda, I\omega\right)
\end{aligned}$$

A lag representation for the expander does not exist, since the output is not wide sense stationary. The cross-power spectral density function is given by

$$\begin{aligned}
S_{xy}^{2D}(\lambda_x,\lambda_y) &= \sum_{m_x=-\infty}^{\infty}\sum_{m_y=-\infty}^{\infty} R_{xy}[m_x,m_y]e^{-j\lambda_x m_x}e^{-j\lambda_y m_y} \\
&= \sum_{m_x=-\infty}^{\infty}\sum_{m_y=-\infty}^{\infty} R_x[m_x,m_y/I]e^{-j\lambda_x m_x}e^{-j\lambda_y m_y} \\
&= \sum_{m_x=-\infty}^{\infty}\sum_{m_y=-\infty}^{\infty} R_x[m_x,m_y/I]e^{-j\lambda_x m_x}e^{-j(I\lambda_y) m_y/I} \\
S_{xy}^{2D}(\lambda_x,\lambda_y) &= S_x^{2D}\left(\lambda_x, \frac{\lambda_y}{L}\right)
\end{aligned}$$

In the time-lag representation, the cross-power spectral density function is defined as

$$
\begin{aligned}
S_{xy}^{2D}(\lambda_x,\omega_y) &= \sum_{m_x=-\infty}^{\infty}\sum_{l_y=-\infty}^{\infty} R_{xy}[m_x;l_y]e^{-j\lambda_x m_x}e^{-j\omega_y l_y} \\
&= \sum_{m_x=-\infty}^{\infty}\sum_{l_y=-\infty}^{\infty} R_x[m_x;(l_y+(I-1)m_x)/I]e^{-j\lambda_x m_x}e^{-j\omega_y l_y} \\
&= \sum_{m_y=-\infty}^{\infty}\sum_{l_y=-\infty}^{\infty} R_x[m_x;(l_y+(I-1)m_x)/I]e^{-j\lambda_x m_x}e^{-j(-(I-1)\omega_y m_x)}e^{-j(I\omega_y)l_y/I}e^{-j(I-1)\omega_y m_x} \\
&= \sum_{m_y=-\infty}^{\infty}\sum_{l_y=-\infty}^{\infty} R_x[m_x;(l_y+(I-1)m_x)/I]e^{-j(\lambda_x-(I-1)\omega_y)m_x}e^{-j(I\omega_y)(l_y+(I-1)m_x)/I} \\
S_{xy}^{2D}(\lambda_x,\omega_y) &= S_x^{2D}\left(\lambda_x-(I-1)\omega_y, I\omega_y\right)
\end{aligned}
$$

C. FILTERS

Filtering is depicted in Fig. A.3 and defined as

$$y[m]=\sum_{r=-\infty}^{\infty} h[r]x[m-r].$$

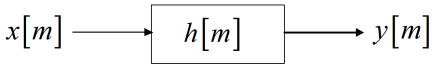


Figure A.3. Filter

The correlation functions for the filter are as shown in Table 2.4 in Chapter II. The output power spectral density function for a filter is

$$
\begin{aligned}
S_y^{2D}(\lambda_1,\lambda_0) &= \sum_{m=-\infty}^{\infty}\sum_{m'=-\infty}^{\infty} R_y[m,m']e^{-j\lambda_1 m}e^{-j\lambda_0 m'} \\
&= \sum_{m=-\infty}^{\infty}\sum_{m'=-\infty}^{\infty} \left(R_x[m,m']*h[m]*h^*[m']\right)e^{-j\lambda_1 m}e^{-j\lambda_0 m'} \\
S_y^{2D}(\lambda_1,\lambda_0) &= S_x^{2D}(\lambda_1,\lambda_0)H(\lambda_1)H^*(\lambda_0)
\end{aligned}
$$

In the time-lag representation, the output power spectral density function is defined as

$$\begin{aligned} S_y^{2D}(\lambda,\omega) &= \sum_{m=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} R_y[m;l]e^{-j\lambda m}e^{-j\omega l} \\ &= \sum_{m=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} \left(R_x[m;l]*h[m]*h^*[-l]\right)e^{-j\lambda m}e^{-j\omega l} \\ S_y^{2D}(\lambda,\omega) &= S_x^{2D}(\lambda,\omega)H(\lambda)H^*(\omega) \end{aligned}$$

or in lag representation only

$$\begin{aligned} S_y(\omega) &= \sum_{l=-\infty}^{\infty} R_y[l]e^{-j\omega l} \\ &= \sum_{l=-\infty}^{\infty} \left(R_x[l]*h[l]*h^*[l]\right)e^{-j\omega l} \\ S_y(\omega) &= S_x(\omega)H(\omega)H^*(\omega) \end{aligned}$$

The cross-power spectral density function is given by

$$\begin{aligned} S_{xy}^{2D}(\lambda_1,\lambda_0) &= \sum_{m=-\infty}^{\infty}\sum_{m'=-\infty}^{\infty} R_{xy}[m,m']e^{-j\lambda_1 m}e^{-j\lambda_0 m'} \\ &= \sum_{m=-\infty}^{\infty}\sum_{m'=-\infty}^{\infty} R_x[m,m']*h^*[m']e^{-j\lambda_1 m}e^{-j\lambda_0 m'} \\ S_{xy}^{2D}(\lambda_1,\lambda_0) &= S_x^{2D}(\lambda_1,\lambda_0)H^*(\lambda_0) \end{aligned}$$

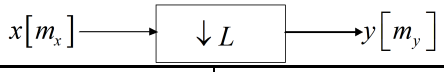
In the time-lag representation, the cross-power spectral density function is defined as

$$\begin{aligned} S_{xy}^{2D}(\lambda,\omega) &= \sum_{m=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} R_{xy}[m;l]e^{-j\lambda m}e^{-j\omega l} \\ &= \sum_{m=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} R_x[m;l]*h^*[-l]e^{-j\lambda m}e^{-j\omega l} \\ S_{xy}^{2D}(\lambda,\omega) &= S_x^{2D}(\lambda,\omega)H^*(\omega) \end{aligned}$$

D. SUMMARY OF CORRELATION AND POWER SPECTRAL DENSITY

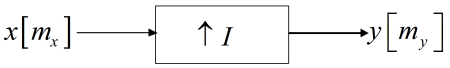
For ease of reference, a summary of the correlation functions and power spectral densities for decimation, expansion and filtering is provided in Tables A.1 through A.3.

Table A.1. Summary of Decimation



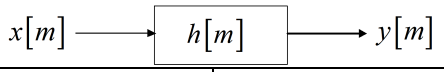
$R_y[m_y, m'_y] = R_x[Lm_y, Lm'_y]$	$S_y^{2D}(\lambda_1, \lambda_0) = S_x^{2D}(\frac{\lambda_1}{L}, \frac{\lambda_0}{L})$
$R_y[m_y; l_y] = R_x[Lm_y; Ll_y]$	$S_y^{2D}(\lambda, \omega) = S_x^{2D}\left(\frac{\lambda}{L}, \frac{\omega}{L}\right)$
$R_y[l_y] = R_x[Ll_y]$	$S_y(\omega) = S_x\left(\frac{\omega}{L}\right)$
$R_{xy}[m_x, m_y] = R_x[m_x, Lm_y]$	$S_{xy}^{2D}(\lambda_x, \lambda_y) = S_x^{2D}\left(\lambda_x, \frac{\lambda_y}{L}\right)$
$R_{xy}[m_x; l_y] = R_x[m_x; Ll_y - (L-1)m_x]$	$S_{xy}^{2D}(\lambda_x, \omega_y) = S_x^{2D}\left(\lambda_x + \frac{L-1}{L}\omega_y, \frac{\omega_y}{L}\right)$

Table A.2. Summary of Expansion



$R_y[m_y, m'_y] = \begin{cases} R_x[m_y/I, m'_y/I] & m_y, m'_y \text{ div } I \\ 0 & \text{otherwise} \end{cases}$	$S_y^{2D}(\lambda_1, \lambda_0) = S_x^{2D}(I\lambda_1, I\lambda_0)$
$R_y[m_y; l_y] = \begin{cases} R_x[m_y/I; l_y/I] & m_y, l_y \text{ div } I \\ 0 & \text{otherwise} \end{cases}$	$S_y^{2D}(\lambda, \omega) = S_x^{2D}(I\lambda, I\omega)$
$R_{xy}[m_x, m_y] = \begin{cases} R_x[m_x, m_y/I] & m_y \text{ div } I \\ 0 & \text{otherwise} \end{cases}$	$S_{xy}^{2D}(\lambda_x, \lambda_y) = S_x^{2D}(\lambda_x, I\lambda_y)$
$R_{xy}[m_x; l_y] = \begin{cases} R_x[m_x; (l_y + (I-1)m_x)/I] & m_x - l_y \text{ div } I \\ 0 & \text{otherwise} \end{cases}$	$S_{xy}^{2D}(\lambda_x, \omega_y) = S_x^{2D}(\lambda_x - (I-1)\omega_y, I\omega_y)$

Table A.3. Summary of Filtering



$R_y[m, m'] = R_x[m, m'] * h[m] * h^*[m']$	$S_y^{2D}(\lambda_1, \lambda_0) = S_x^{2D}(\lambda_1, \lambda_0)H(\lambda_1)H^*(\lambda_0)$
$R_y[l] = R_y[l] * h[l] * h^*[-l]$	$S_y(\omega) = S_x(\omega)H(\omega)H^*(\omega)$
$R_{xy}[m, m'] = R_x[m, m'] * h^*[m']$	$S_{xy}^{2D}(\lambda_1, \lambda_0) = S_x^{2D}(\lambda_1, \lambda_0)H^*(\lambda_0)$
$R_{xy}[m, l] = R_x[m; l] * h^*[-l]$	$S_{xy}^{2D}(\lambda_1, \lambda_0) = S_x^{2D}(\lambda, \omega)H^*(\omega)$

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B. KRONECKER PRODUCT AND REVERSAL NOTATION

A. KRONECKER PRODUCT

This appendix provides the definition of a Kronecker Product and lists some useful relations. A thorough discussion of Kronecker products and matrix calculus is contained in [Ref. 62] and a shorter concise synopsis is contained in [Ref. 63].

Given two matrices $\mathbf{A}$ and $\mathbf{B}$, the Kronecker product $\mathbf{A} \otimes \mathbf{B}$ is defined as

$$\mathbf{A} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \dots & a_{1j}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \dots & a_{2j}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i1}\mathbf{B} & a_{i2}\mathbf{B} & \dots & a_{ij}\mathbf{B} \end{bmatrix}$$

Some basic properties and relations for Kronecker products are provided in Table B.1.

Table B.1. Some Properties of Kronecker Products

(1)	$(\mathbf{A} \otimes \mathbf{B})(\mathbf{D} \otimes \mathbf{G}) = \mathbf{AD} \otimes \mathbf{BG}$
(2)	$(\mathbf{A} + \mathbf{H}) \otimes (\mathbf{B} + \mathbf{R}) = \mathbf{A} \otimes \mathbf{B} + \mathbf{A} \otimes \mathbf{R} + \mathbf{H} \otimes \mathbf{B} + \mathbf{H} \otimes \mathbf{R}$
(3)	$(\mathbf{A} \otimes \mathbf{B})^T = \mathbf{A}^T \otimes \mathbf{B}^T$
(4)	$(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$
(5)	$\mathrm{tr}(\mathbf{A} \otimes \mathbf{B}) = \mathrm{tr}(\mathbf{A}) \cdot \mathrm{tr}(\mathbf{B})$
(6)	$(\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C})$

B. REVERSAL NOTATION

Given a vector $\mathbf{v}$ of length P whose elements are

$$\mathbf{v} = \begin{bmatrix} a_1 & a_2 & \cdots & a_P \end{bmatrix}^T$$

the reversal of the vector, denoted by $\tilde{\mathbf{v}}$, is defined by

$$\tilde{\mathbf{v}} = \begin{bmatrix} a_P & a_{P-1} & \cdots & a_1 \end{bmatrix}^T .$$

Given a $P \times Q$ matrix $\mathbf{A}$ with elements $a_{i,j}$,

$$\mathbf{A} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,Q} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,Q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{P,1} & a_{P,2} & \cdots & a_{P,Q} \end{bmatrix},$$

the reversal of $\mathbf{A}$ is given by

$$\tilde{\mathbf{A}} = \begin{bmatrix} a_{P,Q} & a_{P,Q-1} & \cdots & a_{P,1} \\ a_{P-1,Q} & a_{P-1,Q-1} & \cdots & a_{P-1,1} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1,Q} & a_{1,Q-1} & \cdots & a_{1,1} \end{bmatrix}.$$

A few important properties of vector and matrix reversal are listed in Table B.2.

Table B.2. Properties of Reversal (from [Ref. 18])

	Quantity	Reversal
Matrix product	$\mathbf{AB}$	$\tilde{\mathbf{A}}\tilde{\mathbf{B}}$
Matrix inverse	$\mathbf{A}^{-1}$	$(\tilde{\mathbf{A}})^{-1}$
Matrix conjugate	$\mathbf{A}^{*}$	$(\tilde{\mathbf{A}})^{*}$
Matrix transpose	$\mathbf{A}^{T}$	$(\tilde{\mathbf{A}})^{T}$

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX C. DERIVATION OF PARAMETERS FOR INNOVATIONS FORM OF THE OPTIMAL FILTER

In Chapter III recursive optimal filters were used to present the optimal filter in its *innovations representation*. This appendix contains the derivations related to these optimal filters.

A. REPRESENTING THE OPTIMAL FILTER IN A RECURSIVE FORM

The Wiener-Hopf equations pertaining to an optimal multirate filter with M input channels are given by (3.7). The solution for the filter coefficients can be written as

$$\begin{bmatrix} \mathbf{h}_1^{(k)} \\ \vdots \\ \mathbf{h}_M^{(k)} \end{bmatrix} = \mathbf{C}_M^{(k)^{-1}} \tilde{\mathbf{b}}_{d(M+1)}^{(k)}, \tag{C.1}$$

where

$$\mathbf{C}_M^{(k)} = \begin{bmatrix} \tilde{\mathbf{R}}_{11}^{(k)} & \cdots & \tilde{\mathbf{R}}_{1M}^{(k)} \\ \vdots & \ddots & \vdots \\ \tilde{\mathbf{R}}_{i1}^{(k)*T} & \cdots & \tilde{\mathbf{R}}_{MM}^{(k)} \end{bmatrix} \tag{C.2}$$

and

$$\tilde{\mathbf{b}}_{d(M+1)}^{(k)} = \begin{bmatrix} \tilde{\mathbf{r}}_{d1}^{(k)} \\ \vdots \\ \tilde{\mathbf{r}}_{dM}^{(k)} \end{bmatrix}. \tag{C.3}$$

In general, for the i^{th} signal (C.2) and (C.3) can be defined as

$$\mathbf{C}_i^{(k)} = \begin{bmatrix} \tilde{\mathbf{R}}_{11}^{(k)} & \cdots & \tilde{\mathbf{R}}_{1i}^{(k)} \\ \vdots & \ddots & \vdots \\ \tilde{\mathbf{R}}_{i1}^{(k)*T} & \cdots & \tilde{\mathbf{R}}_{ii}^{(k)} \end{bmatrix} \tag{C.4}$$

and

$$\tilde{\mathbf{b}}_{di}^{(k)} = \begin{bmatrix} \tilde{\mathbf{r}}_{d1}^{(k)} \\ \vdots \\ \tilde{\mathbf{r}}_{d(i-1)}^{(k)} \end{bmatrix}. \tag{C.5}$$

In order to derive a recursive form of the optimal filter, the inverse matrix of $\mathbf{C}_i^{(k)}$ is first written in a recursive form. By partitioning $\mathbf{C}_i^{(k)}$ into

$$\mathbf{C}_i^{(k)} = \begin{bmatrix} \mathbf{C}_{i-1}^{(k)} & \tilde{\mathbf{B}}_{ii}^{(k)} \\ \tilde{\mathbf{B}}_{ii}^{(k)*T} & \tilde{\mathbf{R}}_{ii}^{(k)} \end{bmatrix}, \tag{C.6}$$

where

$$\tilde{\mathbf{B}}_{ij}^{(k)} = \begin{bmatrix} \tilde{\mathbf{R}}_{1j}^{(k)} \\ \vdots \\ \tilde{\mathbf{R}}_{(i-1)j}^{(k)} \end{bmatrix},$$

the recursive form of $\mathbf{C}_i^{(k)^{-1}}$ (see [Ref. 69]) can be expressed as

$$\begin{aligned} \mathbf{C}_i^{(k)^{-1}} &= \begin{bmatrix} \mathbf{C}_{i-1}^{(k)^{-1}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} -\mathbf{G}_i^{(k)} \\ \mathbf{I} \end{bmatrix} \mathbf{E}_i^{(k)^{-1}} \begin{bmatrix} -\mathbf{G}_i^{(k)*T} & \mathbf{I} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{C}_{i-1}^{(k)^{-1}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{G}_i^{(k)}\mathbf{E}_i^{(k)^{-1}}\mathbf{G}_i^{(k)*T} & -\mathbf{G}_i^{(k)}\mathbf{E}_i^{(k)^{-1}} \\ -\mathbf{E}_i^{(k)^{-1}}\mathbf{G}_i^{(k)*T} & \mathbf{E}_i^{(k)^{-1}} \end{bmatrix}, \end{aligned} \tag{C.7}$$

where

$$\mathbf{G}_i^{(k)} = \mathbf{C}_{i-1}^{(k)^{-1}} \tilde{\mathbf{B}}_{ii}^{(k)} \tag{C.8}$$

and

$$\begin{aligned} \mathbf{E}_i^{(k)} &= \tilde{\mathbf{R}}_{ii}^{(k)} - \tilde{\mathbf{B}}_{ii}^{(k)^T} \mathbf{C}_{i-1}^{(k)^{-1}} \tilde{\mathbf{B}}_{ii}^{(k)} \\ &= \tilde{\mathbf{R}}_{ii}^{(k)} - \mathbf{G}_i^{(k)*T} \tilde{\mathbf{B}}_{ii}^{(k)}. \end{aligned} \tag{C.9}$$

Applying the result of (C.7) specifcally to the M^{th} signal by substituting (C.7) into (C.1) leads to

$$\begin{bmatrix} \mathbf{h}_1^{(k)} \\ \vdots \\ \mathbf{h}_M^{(k)} \end{bmatrix} = \left[\begin{bmatrix} \mathbf{C}_{M-1}^{(k)^{-1}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{G}_M^{(k)} \mathbf{E}_M^{(k)^{-1}} \mathbf{G}_M^{(k)*T} & -\mathbf{G}_M^{(k)} \mathbf{E}_M^{(k)^{-1}} \\ -\mathbf{E}_M^{(k)^{-1}} \mathbf{G}_M^{(k)*T} & \mathbf{E}_M^{(k)^{-1}} \end{bmatrix} \right] \begin{bmatrix} \tilde{\mathbf{b}}_{dM}^{(k)} \\ \tilde{\mathbf{r}}_{dM}^{(k)} \end{bmatrix}.$$

Separating the two components of $\mathbf{C}_M^{(k)^{-1}}$ then produces

$$\begin{bmatrix} \mathbf{h}_1^{(k)} \\ \vdots \\ \mathbf{h}_M^{(k)} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_{M-1}^{(k)^{-1}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{b}}_{dM}^{(k)} \\ \tilde{\mathbf{r}}_{dM}^{(k)} \end{bmatrix} + \begin{bmatrix} \mathbf{G}_M^{(k)} \mathbf{E}_M^{(k)^{-1}} \mathbf{G}_M^{(k)*T} & -\mathbf{G}_M^{(k)} \mathbf{E}_M^{(k)^{-1}} \\ -\mathbf{E}_M^{(k)^{-1}} \mathbf{G}_M^{(k)*T} & \mathbf{E}_M^{(k)^{-1}} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{b}}_{dM}^{(k)} \\ \tilde{\mathbf{r}}_{dM}^{(k)} \end{bmatrix}.$$

Extracting the terms associated with $\mathbf{h}_M^{(k)}$, the filter $\mathbf{h}_M^{(k)}$ is found to be equivalent to

$$\begin{aligned} \mathbf{h}_M^{(k)} &= -\mathbf{E}_M^{(k)^{-1}} \mathbf{G}_M^{(k)*T} \tilde{\mathbf{b}}_{dM}^{(k)} + \mathbf{E}_M^{(k)^{-1}} \tilde{\mathbf{r}}_{dM}^{(k)} \\ &= \mathbf{E}_M^{(k)^{-1}} \left(\tilde{\mathbf{r}}_{dM}^{(k)} - \mathbf{G}_M^{(k)*T} \tilde{\mathbf{b}}_{dM}^{(k)} \right) \end{aligned} \tag{C.10}$$

and the rest of the matrices reduce to

$$\begin{bmatrix} \mathbf{h}_1^{(k)} \\ \vdots \\ \mathbf{h}_{M-1}^{(k)} \end{bmatrix} = \mathbf{C}_{M-1}^{(k)^{-1}} \tilde{\mathbf{b}}_{dM}^{(k)} + \begin{bmatrix} \mathbf{G}_M^{(k)} \mathbf{E}_M^{(k)^{-1}} \mathbf{G}_M^{(k)*T} & -\mathbf{G}_M^{(k)} \mathbf{E}_M^{(k)^{-1}} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{b}}_{dM}^{(k)} \\ \tilde{\mathbf{r}}_{dM}^{(k)} \end{bmatrix}$$

or

$$\begin{bmatrix} \mathbf{h}_1^{(k)} \\ \vdots \\ \mathbf{h}_{M-1}^{(k)} \end{bmatrix} = \mathbf{C}_{M-1}^{(k)^{-1}} \tilde{\mathbf{b}}_{dM}^{(k)} + \mathbf{G}_M^{(k)} \mathbf{E}_M^{(k)^{-1}} \mathbf{G}_M^{(k)*T} \tilde{\mathbf{b}}_{dM}^{(k)} - \mathbf{G}_M^{(k)} \mathbf{E}_M^{(k)^{-1}} \tilde{\mathbf{r}}_{dM}^{(k)}.$$

Further manipulation leads to

$$\begin{bmatrix} \mathbf{h}_1^{(k)} \\ \vdots \\ \mathbf{h}_{M-1}^{(k)} \end{bmatrix} = \mathbf{C}_{M-1}^{(k)^{-1}} \tilde{\mathbf{b}}_{dM}^{(k)} - \mathbf{G}_M^{(k)} \mathbf{E}_M^{(k)^{-1}} \left(\tilde{\mathbf{r}}_{dM}^{(k)} - \mathbf{G}_M^{(k)*T} \tilde{\mathbf{b}}_{dM}^{(k)} \right), \tag{C.11}$$

and substituting (C.8) and (C.10) into (C.11) produces

$$\begin{bmatrix} \mathbf{h}_1^{(k)} \\ \vdots \\ \mathbf{h}_{M-1}^{(k)} \end{bmatrix} = \mathbf{C}_{M-1}^{(k)^{-1}} \tilde{\mathbf{b}}_{dM}^{(k)} - \mathbf{C}_{M-1}^{-1} \tilde{\mathbf{B}}_{MM}^{(k)} \mathbf{h}_M^{(k)}. \tag{C.12}$$

Equation (C.12) can also be expressed as

$$\begin{bmatrix} \mathbf{h}_1^{(k)} \\ \vdots \\ \mathbf{h}_{M-1}^{(k)} \end{bmatrix} = \mathbf{C}_{M-1}^{(k)^{-1}} \begin{bmatrix} \tilde{\mathbf{b}}_{d(M-1)}^{(k)} \\ \tilde{\mathbf{r}}_{d(M-1)} \end{bmatrix} - \mathbf{C}_{M-1}^{-1} \begin{bmatrix} \tilde{\mathbf{B}}_{(M-1)M}^{(k)} \\ \tilde{\mathbf{R}}_{(M-1)M}^{(k)} \end{bmatrix} \mathbf{h}_M^{(k)}, \tag{C.13}$$

and substituting (C.7) into (C.13) results in

$$\begin{aligned} \begin{bmatrix} \mathbf{h}_1^{(k)} \\ \vdots \\ \mathbf{h}_{M-1}^{(k)} \end{bmatrix} = & \left(\begin{bmatrix} \mathbf{C}_{M-2}^{(k)^{-1}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{G}_{M-1}^{(k)} \mathbf{E}_{M-1}^{(k)^{-1}} \mathbf{G}_{M-1}^{(k)*T} & -\mathbf{G}_{M-1}^{(k)} \mathbf{E}_{M-1}^{(k)^{-1}} \\ -\mathbf{E}_{M-1}^{(k)^{-1}} \mathbf{G}_{M-1}^{(k)*T} & \mathbf{E}_{M-1}^{(k)^{-1}} \end{bmatrix} \right) \begin{bmatrix} \tilde{\mathbf{b}}_{d(M-1)}^{(k)} \\ \tilde{\mathbf{r}}_{d(M-1)} \end{bmatrix} \\ & - \left(\begin{bmatrix} \mathbf{C}_{M-2}^{(k)^{-1}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{G}_{M-1}^{(k)} \mathbf{E}_{M-1}^{(k)^{-1}} \mathbf{G}_{M-1}^{(k)*T} & -\mathbf{G}_{M-1}^{(k)} \mathbf{E}_{M-1}^{(k)^{-1}} \\ -\mathbf{E}_{M-1}^{(k)^{-1}} \mathbf{G}_{M-1}^{(k)*T} & \mathbf{E}_{M-1}^{(k)^{-1}} \end{bmatrix} \right) \begin{bmatrix} \tilde{\mathbf{B}}_{(M-1)M}^{(k)} \mathbf{h}_M^{(k)} \\ \tilde{\mathbf{R}}_{(M-1)M}^{(k)} \mathbf{h}_M^{(k)} \end{bmatrix} \end{aligned}$$

Extracting the terms associated with $\mathbf{h}_{M-1}^{(k)}$, leads to

$$\begin{aligned} \mathbf{h}_{M-1}^{(k)} = & -\mathbf{E}_{M-1}^{(k)^{-1}} \mathbf{G}_{M-1}^{(k)*T} \tilde{\mathbf{b}}_{d(M-1)}^{(k)} + \mathbf{E}_{M-1}^{(k)^{-1}} \tilde{\mathbf{r}}_{d(M-1)}^{(k)} \\ & + \mathbf{E}_{M-1}^{(k)^{-1}} \mathbf{G}_{M-1}^{(k)*T} \tilde{\mathbf{B}}_{(M-1)M}^{(k)} \mathbf{h}_M^{(k)} - \mathbf{E}_{M-1}^{(k)^{-1}} \tilde{\mathbf{R}}_{(M-1)M}^{(k)} \mathbf{h}_M^{(k)} \end{aligned}$$

This equation can be expressed equivalently as

$$\begin{aligned} \mathbf{h}_{M-1}^{(k)} = & \mathbf{E}_{M-1}^{(k)^{-1}} \left(\tilde{\mathbf{r}}_{d(M-1)}^{(k)} - \mathbf{G}_{M-1}^{(k)*T} \tilde{\mathbf{b}}_{d(M-1)}^{(k)} \right) \\ & - \mathbf{E}_{M-1}^{(k)^{-1}} \left(\tilde{\mathbf{R}}_{(M-1)M}^{(k)} - \mathbf{G}_{M-1}^{(k)*T} \tilde{\mathbf{B}}_{(M-1)M}^{(k)} \right) \mathbf{h}_M^{(k)} \end{aligned} \tag{C.14}$$

The terms for $\mathbf{h}_1^{(k)}$ through $\mathbf{h}_{M-2}^{(k)}$ are then written as

$$\begin{bmatrix} \mathbf{h}_1^{(k)} \\ \vdots \\ \mathbf{h}_{M-2}^{(k)} \end{bmatrix} = \mathbf{C}_{M-2}^{(k)^{-1}} \tilde{\mathbf{b}}_{d(M-1)}^{(k)} + \begin{bmatrix} \mathbf{G}_{M-1}^{(k)} \mathbf{E}_{M-1}^{(k)^{-1}} \mathbf{G}_{M-1}^{(k)*T} & -\mathbf{G}_{M-1}^{(k)} \mathbf{E}_{M-1}^{(k)^{-1}} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{b}}_{d(M-1)}^{(k)} \\ \tilde{\mathbf{r}}_{d(M-1)}^{(k)} \end{bmatrix}$$
$$- \mathbf{C}_{M-2}^{(k)^{-1}} \tilde{\mathbf{B}}_{(M-1)M}^{(k)} \mathbf{h}_M^{(k)} - \begin{bmatrix} \mathbf{G}_{M-1}^{(k)} \mathbf{E}_{M-1}^{(k)^{-1}} \mathbf{G}_{M-1}^{(k)*T} & -\mathbf{G}_{M-1}^{(k)} \mathbf{E}_{M-1}^{(k)^{-1}} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{B}}_{(M-1)M}^{(k)} \mathbf{h}_M^{(k)} \\ \tilde{\mathbf{R}}_{(M-1)M}^{(k)} \mathbf{h}_M^{(k)} \end{bmatrix}$$

This equation can be rearranged, producing

$$\begin{aligned} \begin{bmatrix} \mathbf{h}_1^{(k)} \\ \vdots \\ \mathbf{h}_{M-2}^{(k)} \end{bmatrix} = \mathbf{C}_{M-2}^{(k)^{-1}} \tilde{\mathbf{b}}_{d(M-1)}^{(k)} - \mathbf{G}_{M-1}^{(k)} \mathbf{E}_{M-1}^{(k)^{-1}} \left(\tilde{\mathbf{r}}_{d(M-1)}^{(k)} - \mathbf{G}_{M-1}^{(k)*T} \tilde{\mathbf{b}}_{d(M-1)}^{(k)} \right) \\ + \mathbf{G}_{M-1}^{(k)} \mathbf{E}_{M-1}^{(k)^{-1}} \left(\tilde{\mathbf{R}}_{(M-1)M}^{(k)} - \mathbf{G}_{M-1}^{(k)*T} \tilde{\mathbf{B}}_{(M-1)M}^{(k)} \right) \mathbf{h}_M^{(k)} \\ - \mathbf{C}_{M-2}^{(k)^{-1}} \tilde{\mathbf{B}}_{(M-1)M}^{(k)} \mathbf{h}_M^{(k)}. \end{aligned} \tag{C.15}$$

Now substituting (C.8) and (C.14) into (C.15) yields

$$\begin{bmatrix} \mathbf{h}_1^{(k)} \\ \vdots \\ \mathbf{h}_{M-2}^{(k)} \end{bmatrix} = \mathbf{C}_{M-2}^{(k)^{-1}} \tilde{\mathbf{b}}_{d(M-1)}^{(k)} - \mathbf{C}_{M-2}^{(k)^{-1}} \tilde{\mathbf{B}}_{(M-1)(M-1)}^{(k)} \mathbf{h}_{M-1}^{(k)} - \mathbf{C}_{M-2}^{(k)^{-1}} \tilde{\mathbf{B}}_{(M-1)M}^{(k)} \mathbf{h}_M^{(k)}. \tag{C.16}$$

By examining the pattern of how (C.1) is separated into (C.10) and (C.12) and how (C.12) is separated into (C.14) and (C.16), the following recursions can be observed

$$\begin{aligned} \mathbf{h}_i^{(k)} &= \mathbf{E}_i^{(k)^{-1}} \left(\tilde{\mathbf{r}}_{di}^{(k)} - \mathbf{G}_i^{(k)*T} \tilde{\mathbf{b}}_{di}^{(k)} \right) \\ &\quad - \sum_{j=i+1}^{M} \mathbf{E}_i^{(k)^{-1}} \left(\tilde{\mathbf{R}}_{ij}^{(k)} - \mathbf{G}_i^{(k)*T} \tilde{\mathbf{B}}_{ij}^{(k)} \right) \mathbf{h}_j^{(k)} \qquad (1 \le i \le M-1) \end{aligned} \tag{C.17}$$

$$\mathbf{h}_M^{(k)} = \mathbf{E}_M^{(k)^{-1}} \left(\tilde{\mathbf{r}}_{dM}^{(k)} - \mathbf{G}_M^{(k)*T} \tilde{\mathbf{b}}_{dM}^{(k)} \right) \tag{C.18}$$

where the matrices $\mathbf{G}_i^{(k)}$ and $\mathbf{E}_i^{(k)}$ are given by

$$\mathbf{G}_i^{(k)} = \begin{cases} 0 & i = 1 \\ \begin{bmatrix} \tilde{\mathbf{R}}_{11}^{(k)} & \tilde{\mathbf{R}}_{12}^{(k)} & \cdots & \tilde{\mathbf{R}}_{1(i-1)}^{(k)} \\ \tilde{\mathbf{R}}_{12}^{(k)*T} & \tilde{\mathbf{R}}_{2}^{(k)} & \cdots & \tilde{\mathbf{R}}_{2(i-1)}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\mathbf{R}}_{1(i-1)}^{(k)*T} & \tilde{\mathbf{R}}_{2(i-1)}^{(k)*T} & \cdots & \tilde{\mathbf{R}}_{(i-1)(i-1)}^{(k)} \end{bmatrix}^{-1} \tilde{\mathbf{B}}_{ii}^{(k)} & 1 < i \leq M \end{cases}$$

and

$$\mathbf{E}_i^{(k)} = \begin{cases} \tilde{\mathbf{R}}_{11}^{(k)} & i = 1 \\ \tilde{\mathbf{R}}_{ii}^{(k)} - \tilde{\mathbf{G}}_i^{(k)*T} \mathbf{B}_{ii}^{(k)} & 1 < i \leq M \end{cases}.$$

The filter in (C.17) and (C.18) can be further simplified by defining

$$\mathbf{H}_{ij}^{(k)} = \mathbf{E}_i^{(k)^{-1}} \left(\tilde{\mathbf{R}}_{ij}^{(k)} - \mathbf{G}_i^{(k)*T} \tilde{\mathbf{B}}_{ij}^{(k)} \right) \tag{C.19}$$

$$\mathbf{h}_i^{'(k)} = \mathbf{E}_i^{(k)^{-1}} \left(\tilde{\mathbf{r}}_{di}^{(k)} - \mathbf{G}_i^{(k)*T} \tilde{\mathbf{b}}_{di}^{(k)} \right) \tag{C.20}$$

and substituting these terms into (C.17) and (C.18)

$$\mathbf{h}_i^{(k)} = \mathbf{h}_i^{'(k)} - \sum_{j=i+1}^{M} \mathbf{H}_{ij}^{(k)} \mathbf{h}_j^{(k)} \qquad (1 \leq i \leq M-1) \tag{C.21}$$

$$\mathbf{h}_M^{(k)} = \mathbf{h}_M^{'(k)} \tag{C.22}$$

The filter coefficients of (C.21) for an M-signal system can be expressed in terms of the cross terms $\mathbf{H}_{ij}^{(k)}$ and orthogonal filters $\mathbf{h}_i^{'(k)}$ only, which is useful for development of the recursive error variance. Beginning with the general equation for the filters for signals 1 to $M-1$, (C.21), but using the index i_1 instead of j for the summation

$$\mathbf{h}_i^{(k)} = \mathbf{h}_i^{'(k)} + \sum_{i_1=i+1}^{M} \left(-\mathbf{H}_{ii_1}^{(k)} \right) \mathbf{h}_{i_1}^{(k)} \quad (1 \leq i \leq M-1) \tag{C.23}$$

the filter for signal i can be expanded by substituting (C.21) for the filter $\mathbf{h}_{i_1}^{(k)}$

$$\mathbf{h}_i^{(k)} = \mathbf{h}_i^{'(k)} + \sum_{i_1=i+1}^{M} \left(-\mathbf{H}_{ii_1}^{(k)} \right) \left(\mathbf{h}_{i_1}^{'(k)} + \sum_{i_2=i_1+1}^{M} \left(-\mathbf{H}_{i_1 i_2}^{(k)} \right) \mathbf{h}_{i_2}^{(k)} \right) \quad (1 \leq i \leq M-1). \tag{C.24}$$

Separating the expanded term yields

$$\mathbf{h}_i^{(k)} = \mathbf{h}_i^{'(k)} + \sum_{i_1=i+1}^{M} \left(-\mathbf{H}_{ii_1}^{(k)}\right) \mathbf{h}_{i_1}^{'(k)} + \sum_{i_1=i+1}^{M} \sum_{i_2=i_1+1}^{M} \left(-\mathbf{H}_{ii_1}^{(k)}\right) \left(-\mathbf{H}_{i_1 i_2}^{(k)}\right) \mathbf{h}_{i_2}^{(k)} \quad (1 \le i \le M-1). \tag{C.25}$$

Substituting (C.21) into the third term leads to

$$\mathbf{h}_i^{(k)} = \mathbf{h}_i^{'(k)} + \sum_{i_1=i+1}^{M} \left(-\mathbf{H}_{ii_1}^{(k)}\right) \mathbf{h}_{i_1}^{'(k)} + \sum_{i_1=i+1}^{M} \sum_{i_2=i_1+1}^{M} \left(-\mathbf{H}_{ii_1}^{(k)}\right) \left(-\mathbf{H}_{i_1 i_2}^{(k)}\right) \left(\mathbf{h}_{i_2}^{'(k)} + \sum_{i_3=i_2+1}^{M} \left(-\mathbf{H}_{i_2 i_3}^{(k)}\right) \mathbf{h}_{i_3}^{(k)}\right) \quad (1 \le i \le M-1) \tag{C.26}$$

Separating the last term on the right-hand side of the equation yields

$$\mathbf{h}_i^{(k)} = \mathbf{h}_i^{'(k)} + \sum_{i_1=i+1}^{M} \left(-\mathbf{H}_{ii_1}^{(k)}\right) \mathbf{h}_{i_1}^{'(k)} + \sum_{i_1=i+1}^{M} \sum_{i_2=i_1+1}^{M} \left(-\mathbf{H}_{ii_1}^{(k)}\right) \left(-\mathbf{H}_{i_1 i_2}^{(k)}\right) \mathbf{h}_{i_2}^{'(k)} + \sum_{i_1=i+1}^{M} \sum_{i_2=i_1+1}^{M} \sum_{i_3=i_2+1}^{M} \left(-\mathbf{H}_{ii_1}^{(k)}\right) \left(-\mathbf{H}_{i_1 i_2}^{(k)}\right) \left(-\mathbf{H}_{i_2 i_3}^{(k)}\right) \mathbf{h}_{i_3}^{(k)} \quad (1 \le i \le M-1) \tag{C.27}$$

If this recursion is continued, the final result becomes

$$\mathbf{h}_i^{(k)} = \mathbf{h}_i^{'(k)} + \sum_{i_1=i+1}^{M} \left(-\mathbf{H}_{ii_1}^{(k)}\right) \mathbf{h}_{i_1}^{'(k)} + \sum_{i_1=i+1}^{M} \sum_{i_2=i_1+1}^{M} \left(-\mathbf{H}_{ii_1}^{(k)}\right) \left(-\mathbf{H}_{i_1 i_2}^{(k)}\right) \mathbf{h}_{i_2}^{'(k)} + \cdots + \sum_{i_1=i+1}^{M} \sum_{i_2=i_1+1}^{M} \cdots \sum_{i_{M-1}=i_{M-2}+1}^{M} \sum_{i_M=i_{M-1}+1}^{M} \left(-\mathbf{H}_{ii_1}^{(k)}\right) \left(-\mathbf{H}_{i_1 i_2}^{(k)}\right) \cdots \left(-\mathbf{H}_{i_{M-2} i_{M-1}}^{(k)}\right) \left(-\mathbf{H}_{i_{M-1} i_M}^{(k)}\right) \mathbf{h}_{i_M}^{'(k)} \quad (1 \le i \le M-1). \tag{C.28}$$

The result of (C.28) can be used to also develop the recursive form of the error variance.

B. EXPRESSION FOR THE ERROR VARIANCE

Beginning with the basic definition of the error variance

$$\sigma_k^2 = R_d(0) - \sum_{i=1}^{M} \tilde{\mathbf{r}}_{di}^{(k)\,T} \mathbf{h}_i^{(k)*}, \tag{C.29}$$

the summation can be separated into

$$\sigma_k^2 = R_d(0) - \sum_{i=1}^{M-1} \tilde{\mathbf{r}}_{di}^{(k)\,T} \mathbf{h}_i^{(k)*} - \tilde{\mathbf{r}}_{dM}^{(k)\,T} \mathbf{h}_M^{(k)*}. \tag{C.30}$$

Substituting (C.28) into (C.30) for the filter term representing the filters for signals 1 through $M-1$ leads to

$$\begin{aligned}
\sigma_k^2 = R_d(0) - \sum_{i=1}^{M-1} \tilde{\mathbf{r}}_{di}^{(k)\,T} \times \\
\left(\mathbf{h}_i^{'(k)} + \sum_{i_1=i+1}^{M} \left(-\mathbf{H}_{ii_1}^{(k)}\right) \mathbf{h}_{i_1}^{'(k)} + \sum_{i_1=i+1}^{M} \sum_{i_2=i_1+1}^{M} \left(-\mathbf{H}_{ii_1}^{(k)}\right) \left(-\mathbf{H}_{i_1 i_2}^{(k)}\right) \mathbf{h}_{i_2}^{'(k)} + \cdots \right. \\
+ \sum_{i_1=i+1}^{M} \sum_{i_2=i_1+1}^{M} \cdots \sum_{i_{M-1}=i_{M-2}+1}^{M} \sum_{i_M=i_{M-1}+1}^{M} \left(-\mathbf{H}_{ii_1}^{(k)}\right) \times \\
\left. \left(-\mathbf{H}_{i_1 i_2}^{(k)}\right) \cdots \left(-\mathbf{H}_{i_{M-2} i_{M-1}}^{(k)}\right) \left(-\mathbf{H}_{i_{M-1} i_M}^{(k)}\right) \mathbf{h}_{i_M}^{'(k)} \right)^* \\
- \tilde{\mathbf{r}}_{dM}^{(k)\,T} \mathbf{h}_M^{(k)*}.
\end{aligned} \tag{C.31}$$

The multiple summation term can be separated into two parts leading to

$$\begin{aligned}
\sigma_k^2 = R_d(0) - \sum_{i=1}^{M-1} \tilde{\mathbf{r}}_{di}^{(k)\,T} \times \\
\left(\mathbf{h}_i^{'(k)} + \sum_{i_1=i+1}^{M-1} \left(-\mathbf{H}_{ii_1}^{(k)}\right) \mathbf{h}_{i_1}^{'(k)} + \sum_{i_1=i+1}^{M-1} \sum_{i_2=i_1+1}^{M-1} \left(-\mathbf{H}_{ii_1}^{(k)}\right) \left(-\mathbf{H}_{i_1 i_2}^{(k)}\right) \mathbf{h}_{i_2}^{'(k)} + \cdots \right. \\
+ \sum_{i_1=i+1}^{M-1} \sum_{i_2=i_1+1}^{M-1} \cdots \sum_{i_{M-1}=i_{M-2}+1}^{M-1} \sum_{i_M=i_{M-1}+1}^{M-1} \left(-\mathbf{H}_{ii_1}^{(k)}\right) \times \\
\left. \left(-\mathbf{H}_{i_1 i_2}^{(k)}\right) \cdots \left(-\mathbf{H}_{i_{M-2} i_{M-1}}^{(k)}\right) \mathbf{h}_{i_{M-1}}^{'(k)} \right)^* \\
- \sum_{i=1}^{M-1} \tilde{\mathbf{r}}_{dM}^{(k)\,T} \left(\left(-\mathbf{H}_{iM}^{(k)}\right) \mathbf{h}_{i_M}^{'(k)} + \sum_{i_1=i+1}^{M} \left(-\mathbf{H}_{ii_1}^{(k)}\right) \left(-\mathbf{H}_{i_1 M}^{(k)}\right) \mathbf{h}_M^{'(k)} + \cdots \right. \\
+ \sum_{i_1=i+1}^{M} \sum_{i_2=i_1+1}^{M} \cdots \sum_{i_{M-1}=i_{M-2}+1}^{M} \left(-\mathbf{H}_{ii_1}^{(k)}\right) \times \\
\left. \left(-\mathbf{H}_{i_1 i_2}^{(k)}\right) \cdots \left(-\mathbf{H}_{i_{M-2} i_{M-1}}^{(k)}\right) \left(-\mathbf{H}_{i_{M-1} M}^{(k)}\right) \mathbf{h}_{i_M}^{'(k)} \right)^* \\
- \tilde{\mathbf{r}}_{dM}^{(k)\,T} \mathbf{h}_M^{(k)*}. \quad \text{(C.32)}
\end{aligned}$$

The first two terms on the right-hand side is the error variance for an $(M-1)$-signal system. Thus defining $\sigma_{k,M}^2 = \sigma_k^2$ and representing the error variance for the $(M-1)$-signal system as $\sigma_{k,M-1}^2$ leads to

$$\begin{aligned}
\sigma_{k,M}^2 = \sigma_{k,M-1}^2 - \tilde{\mathbf{r}}_{dM}^{(k)\,T} \mathbf{h}_M^{(k)*} - \sum_{i=1}^{M-1} \tilde{\mathbf{r}}_{dM}^{(k)\,T} \left(-\mathbf{H}_{iM}^{(k)*}\right) \mathbf{h}_{i_M}^{'(k)*} \\
- \sum_{i=1}^{M-1} \sum_{i_1=i+1}^{M} \tilde{\mathbf{r}}_{dM}^{(k)\,T} \left(-\mathbf{H}_{ii_1}^{(k)*}\right) \left(-\mathbf{H}_{i_1 M}^{(k)*}\right) \mathbf{h}_M^{'(k)*} - \cdots \\
- \sum_{i=1}^{M-1} \sum_{i_1=i+1}^{M} \sum_{i_2=i_1+1}^{M} \cdots \sum_{i_{M-1}=i_{M-2}+1}^{M} \tilde{\mathbf{r}}_{dM}^{(k)\,T} \left(-\mathbf{H}_{ii_1}^{(k)*}\right) \times \\
\left(-\mathbf{H}_{i_1 i_2}^{(k)*}\right) \cdots \left(-\mathbf{H}_{i_{M-2} i_{M-1}}^{(k)*}\right) \left(-\mathbf{H}_{i_{M-1} M}^{(k)*}\right) \mathbf{h}_{i_M}^{'(k)*}. \quad \text{(C.33)}
\end{aligned}$$

C. SUMMARY

For clarity and ease of reference, the results for the innovations form of the optimal filter are summarized here. Given an M-signal multirate system, the solution to the Wiener-Hopf equations is given by

$$\begin{bmatrix} \mathbf{h}_1^{(k)} \\ \vdots \\ \mathbf{h}_M^{(k)} \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{R}}_{11}^{(k)} & \cdots & \tilde{\mathbf{R}}_{1M}^{(k)} \\ \vdots & \ddots & \vdots \\ \tilde{\mathbf{R}}_{1M}^{(k)*T} & \cdots & \tilde{\mathbf{R}}_{MM}^{(k)} \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\mathbf{r}}_{d1}^{(k)} \\ \vdots \\ \tilde{\mathbf{r}}_{dM}^{(k)} \end{bmatrix}.$$

The following terms are defined for use in the filter equations

$$\tilde{\mathbf{B}}_{ij}^{(k)} = \begin{bmatrix} \tilde{\mathbf{R}}_{1j}^{(k)} \\ \vdots \\ \tilde{\mathbf{R}}_{(i-1)j}^{(k)} \end{bmatrix} \qquad \tilde{\mathbf{b}}_{di}^{(k)} = \begin{bmatrix} \tilde{\mathbf{r}}_{d1}^{(k)} \\ \vdots \\ \tilde{\mathbf{r}}_{d(i-1)}^{(k)} \end{bmatrix},$$

and

$$\mathbf{G}_i^{(k)} = \begin{cases} 0 & i = 1 \\ \begin{bmatrix} \tilde{\mathbf{R}}_{11}^{(k)} & \tilde{\mathbf{R}}_{12}^{(k)} & \cdots & \tilde{\mathbf{R}}_{1(i-1)}^{(k)} \\ \tilde{\mathbf{R}}_{12}^{(k)*T} & \tilde{\mathbf{R}}_{2}^{(k)} & \cdots & \tilde{\mathbf{R}}_{2(i-1)}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\mathbf{R}}_{1(i-1)}^{(k)*T} & \tilde{\mathbf{R}}_{2(i-1)}^{(k)*T} & \cdots & \tilde{\mathbf{R}}_{(i-1)(i-1)}^{(k)} \end{bmatrix}^{-1} \tilde{\mathbf{B}}_{ii}^{(k)} & 1 < i \leq M \end{cases}$$

and

$$\mathbf{E}_i^{(k)} = \begin{cases} \tilde{\mathbf{R}}_{11}^{(k)} & i = 1 \\ \tilde{\mathbf{R}}_{ii}^{(k)} - \tilde{\mathbf{G}}_i^{(k)*T} \mathbf{B}_{ii}^{(k)} & 1 < i \leq M \end{cases}.$$

The equations for the filter coefficient in the innovations representation of the filter are then summarized in Table C.1.

Table C.1. Innovations Form of the Optimal Filter

$$\mathbf{H}_{ij}^{(k)} = \mathbf{E}_i^{(k)^{-1}} \left(\tilde{\mathbf{R}}_{ij}^{(k)} - \mathbf{G}_i^{(k)*T} \tilde{\mathbf{B}}_{ij}^{(k)} \right)$$

$$\mathbf{h}_i^{'(k)} = \mathbf{E}_i^{(k)^{-1}} \left(\tilde{\mathbf{r}}_{di}^{(k)} - \mathbf{G}_i^{(k)*T} \tilde{\mathbf{b}}_{di}^{(k)} \right)$$

$$\mathbf{h}_i^{(k)} = \mathbf{h}_i^{'(k)} - \sum_{j=i+1}^{M} \mathbf{H}_{ij}^{(k)} \mathbf{h}_j^{(k)} \qquad (1 \le i \le M-1)$$

$$\mathbf{h}_M^{(k)} = \mathbf{h}_M^{'(k)}$$

$$\begin{aligned}
\sigma_{k,M}^2 = \sigma_{k,M-1}^2 &- \tilde{\mathbf{r}}_{dM}^{(k)T} \mathbf{h}_M^{(k)*} - \sum_{i=1}^{M-1} \tilde{\mathbf{r}}_{dM}^{(k)T} \left(-\mathbf{H}_{iM}^{(k)*} \right) \mathbf{h}_{i_M}^{'(k)*} \\
&- \sum_{i=1}^{M-1} \sum_{i_1=i+1}^{M} \tilde{\mathbf{r}}_{dM}^{(k)T} \left(-\mathbf{H}_{ii_1}^{(k)*} \right) \left(-\mathbf{H}_{i_1 M}^{(k)*} \right) \mathbf{h}_M^{'(k)*} - \cdots \\
&- \sum_{i=1}^{M-1} \sum_{i_1=i+1}^{M} \sum_{i_2=i_1+1}^{M} \cdots \sum_{i_{M-1}=i_{M-2}+1}^{M} \tilde{\mathbf{r}}_{dM}^{(k)T} \left(-\mathbf{H}_{ii_1}^{(k)*} \right) \times \\
&\left(-\mathbf{H}_{i_1 i_2}^{(k)*} \right) \cdots \left(-\mathbf{H}_{i_{M-2} i_{M-1}}^{(k)*} \right) \left(-\mathbf{H}_{i_{M-1} M}^{(k)*} \right) \mathbf{h}_{i_M}^{'(k)*}
\end{aligned}$$

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX D. MULTICHANNEL LEVINSON RECURSION

This appendix briefly discusses the procedure known as the multichannel Levinson algorithm, (also called the Levinson-Wiggins-Robinson (LWR) algorithm). For more detailed discussions of this topic see [Ref. 64, 70, 71]. The multichannel Levinson algorithm is a method of solving the Normal equations for multichannel systems in a recursive fashion. This algorithm is an extension of the Levinson recursion for single-channel systems.

A. THE MULTICHANNEL SYSTEM

In order to properly describe the multichannel Levinson algorithm, it is necessary to first consider a multichannel random process and its correlation function. Let the random process $\mathbf{x}[n]$ have the form

$$\mathbf{x}[n] = \begin{bmatrix} x_1[n] \\ x_2[n] \\ \vdots \\ x_C[n] \end{bmatrix},$$

where C is the number of channels. The correlation function associated with $\mathbf{x}[n]$ is given by

$$\mathbf{R}_{\mathbf{x}}[l] = \mathcal{E}\{\mathbf{x}[n]\mathbf{x}^{*T}[n-l]\}.$$

where $\mathbf{R}_{\mathbf{x}}[l]$ is a $C \times C$ matrix.

The multichannel Levinson algorithm solves for the filter coefficients and prediction error covariance matrices for linear prediction of the multichannel random

process. In so doing, it finds the parameters for both the forward and backward prediction of the multichannel random process.

B. GENERAL FORM OF THE ALGORITHM

To initialize the recursion, the following variables are defined,

$$\mathbf{R}_0 = \mathbf{R}_{\mathbf{x}}[1] \qquad \mathbf{A}_0^{'} = \mathbf{B}_0^{'} = \mathbf{I}_{C\times C} \qquad \Sigma_0 = \Sigma_0^b = \mathbf{R}_{\mathbf{x}}[0],$$

where $\mathbf{I}_{C\times C}$ is an identity matrix of size $C \times C$. In addition, a cumulative correlation matrix is defined

$$\mathbf{R}_p = \begin{bmatrix} \mathbf{R}_{\mathbf{x}}[p] \\ \mathbf{R}_{\mathbf{x}}[p-1] \\ \vdots \\ \mathbf{R}_{\mathbf{x}}[1] \end{bmatrix}$$

where p is between 1 and the desired order. Throughout this algorithm, the variables with the superscript b refer to the backward prediction parameters and the other variables refer to the forward prediction parameters.

Step 1. Calculate the reflection coefficients, Γ_p and Γ_p^b

$$\Delta_p = \begin{bmatrix} \mathbf{R}_1^T & \mathbf{R}_2^T & \cdots & \mathbf{R}_p^T \end{bmatrix} \tilde{\mathbf{A}}_{p-1}^{'} = \left[\begin{bmatrix} \mathbf{R}_1 & \mathbf{R}_2 & \cdots & \mathbf{R}_p \end{bmatrix} \tilde{\mathbf{B}}_{p-1}^{'} \right]^T$$

$$\Gamma_p = \Sigma_{p-1}^{b^{-1}} \Delta_p$$

and

$$\Gamma_p^b = \Sigma_{p-1}^{-1} \Delta_p^T$$

Step 2. Calculate the p^{th} order prediction coefficients, $\mathbf{A}_p^{'}$ and $\mathbf{B}_p^{'}$

$$\mathbf{A}_p^{'} = \begin{bmatrix} \mathbf{A}_{p-1}^{'} \\ \text{---} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{0} \\ \text{---} \\ \tilde{\mathbf{B}}_{p-1}^{'} \end{bmatrix} \Gamma_p$$

and

$$\mathbf{B}_p^{'} = \begin{bmatrix} \mathbf{B}_{p-1}^{'} \\ \hline \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{0} \\ \hline \\ \tilde{\mathbf{A}}_{p-1}^{'} \end{bmatrix} \Gamma_p^b$$

Step 3. Calculate the p^{th} order Error Variance coefficients, Σ_p and Σ_p^b

$$\Sigma_p = \Sigma_{p-1} - \Gamma_p^T \Sigma_{p-1}^b \Gamma_p$$

and

$$\Sigma_p^b = \Sigma_{p-1}^b - \Gamma_p^{bT} \Sigma_{p-1} \Gamma_p^b$$

These steps are repeated until the desired order has been reached. At that point, the prediction coefficients, error variances and reflection coefficients for all orders up through the desired order have been calculated.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX E. A GENERALIZED MATRIX TRIANGULAR FACTORIZATION ALGORITHM

In order to solve (4.41) for the transform matrix $\mathbf{A}_0$ when developing a solution to the multirate linear prediction problem in Chapter IV, a generalized UL factorization was developed. The generalized UL factorization presented in this appendix is modeled after the standard LU factorization algorithm that can be found in many linear algebra books, such as [Ref. 65]; however, when factored, the left matrix is upper triangular and the right matrix is lower triangular. The algorithm is called "generalized" since it allows for a block UL factorization using different size blocks. It thus provides for Gaussian elimination of multiple rows at the same time. This is important to the multirate linear prediction solution, since eliminating multiple rows at the same time is necessary when dealing with joint observations of multiple signals.

A. LU TRIANGULAR FACTORIZATION REVIEW

This section briefly reviews the algorithm for computing the basic LU triangular factorization of a matrix. For a more detailed discussion of this methodology, refer to any elementary linear algebra textbook, e.g, [Ref. 65].

Given an $m \times n$ matrix $\mathbf{A}$,

$$\mathbf{A} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m,1} & a_{i,2} & \cdots & a_{m,n} \end{bmatrix},$$

it is desired to decompose the matrix into upper and lower triangular matrices, such that

$$\mathbf{A} = \mathbf{LU},$$

where **L** is the lower triangular matrix with 1's on the main diagonal and **U** is the upper triangular matrix. The procedure for this method is as follows:

Step 1: For $i = 1, 2, \ldots, m-1$, eliminate row i from rows $i+1$ to m:

a. Calculate row multipliers for rows $i+1$ to m

$$M_{k,i} = a_{k,i}/a_{i,i} \ , \quad k = i+1, i+2, \ldots, m$$

b. Subtract row i from rows $i+1$ to m

$$a_{k,i} = a_{k,i} - M_{k,i} * a_{i,i} = 0 \ , \qquad k = i+1, i+2, \ldots, m$$
$$a_{k,l} = a_{k,l} - M_{k,i} * a_{i,l} \qquad , \quad k, l = i+1, i+2, \ldots, m$$

After eliminating i rows, the matrix has the form:

$$\mathbf{A}_i = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \cdots & a_{1,i} & a_{1,i+1} & \cdots & a_{1,n} \\ 0 & a_{2,2} & a_{2,3} & \cdots & a_{2,i} & a_{2,i+1} & \cdots & a_{2,n} \\ 0 & 0 & a_{3,3} & \cdots & a_{3,i} & a_{3,i+1} & \cdots & a_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{i,i} & a_{i,i+1} & \cdots & a_{i,n} \\ 0 & 0 & 0 & \cdots & a_{i+1,i} & a_{i+1,i+1} & \cdots & a_{i+1,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \cdots \\ 0 & 0 & 0 & \cdots & a_{m,i} & a_{m,i+1} & \cdots & a_{m,n} \end{bmatrix}$$

Repeat step 1 until matrix **A** is upper triangular. This is the matrix **U**.

$$\mathbf{U} = \mathbf{A}_m = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \cdots & a_{1,i} & a_{1,i+1} & \cdots & a_{1,n} \\ 0 & a_{2,2} & a_{2,3} & \cdots & a_{2,i} & a_{2,i+1} & \cdots & a_{2,n} \\ 0 & 0 & a_{3,3} & \cdots & a_{3,i} & a_{3,i+1} & \cdots & a_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{i,i} & a_{i,i+1} & \cdots & a_{i,n} \\ 0 & 0 & 0 & \cdots & 0 & a_{i+1,i+1} & \cdots & a_{i+1,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \cdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & a_{m,n} \end{bmatrix}$$

Step 2: The lower triangular matrix can be formed as

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ M_{2,1} & 1 & 0 & \cdots & 0 \\ M_{3,1} & M_{3,2} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ M_{m,1} & M_{m,2} & M_{m,3} & \cdots & 1 \end{bmatrix},$$

where $M_{i,j}$ are defined above.

B. A GENERALIZED UL TRIANGULAR FACTORIZATION ALGORITHM

The basic LU triangular factorization algorithm can be modified and extended to a generalized UL triangular factorization algorithm. This new procedure requires the defining and use of interim variables to transform an original matrix $\mathbf{A}$.

Given an $m \times n$ matrix $\mathbf{A}$,

$$\mathbf{A} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m,1} & a_{i,2} & \cdots & a_{m,n} \end{bmatrix},$$

it is desired to decompose the matrix into generalized upper and lower triangular matrices, such that

$$\mathbf{A} = \mathbf{UL},$$

where $\mathbf{U}$ is a block upper triangular matrix and $\mathbf{L}$ is a block lower triangular matrix. The blocks are not all of the same size, but the sizes of the desired blocks are known (or given).

The procedure for this method is as follows (starting with the *last column*),

Step 1: Eliminate columns $i-r+1$ through i from columns 1 through $i-r$, where r is the number of columns to be eliminated at the same time and i is the index of the last of these columns.

a. Define the following vectors and matrices

$$\mathbf{A}_{(i-r+1)i,(i-r+1)i} = \begin{bmatrix} a_{i-r+1,i-r+1} & \cdots & a_{i-r+1,i} \\ \vdots & \ddots & \vdots \\ a_{i,i-r+1} & \cdots & a_{i,i} \end{bmatrix}$$

$$\mathbf{a}_{(i-r+1)i,k} = \begin{bmatrix} a_{i-r+1,k} \\ \vdots \\ a_{i,k} \end{bmatrix}, \ k = 1, 2, \ldots, i-r$$

$$\mathbf{a}_{l,(i-r+1)i} = \begin{bmatrix} a_{l,i-r+1} & \cdots & a_{l,i} \end{bmatrix}, \ l = 1, 2, \ldots, i-r$$

b. Calculate column multipliers for columns 1 to $i-r$

$$M_{(i-r+1)i,k} = \mathbf{A}^{-1}_{(i-r+1)i,(i-r+1)i}\mathbf{a}_{(i-r+1)i,k}, \ k = 1, 2, \ldots, i-r$$

c. Subtract columns $i-r+1$ through i from columnss 1 through $i-r$

$$\mathbf{a}_{(i-r+1)i,k} = \mathbf{a}_{(i-r+1)i,k} - \mathbf{A}_{(i-r+1)i,(i-r+1)i}M_{(i-r+1)i,k} = \mathbf{0} \quad , \quad k = 1, 2, \ldots, i-r$$

$$a_{l,k} = a_{l,k} - \mathbf{a}_{l,(i-r+1)i}M_{(i-r+1)i,k} \quad , \ k, l = 1, 2, \ldots, i-r$$

At the i^{th} iteration, the matrix has the form:

$$\mathbf{A}_i = \begin{bmatrix} a_{1,1} & \cdots & a_{1,i-r} & & \mathbf{a}_{1,(i-r+1)i} & & \cdots & a_{1,n-2} & a_{1,n-1} & a_{1,n} \\ \vdots & \ddots & \vdots & & \vdots & & \ddots & \vdots & \vdots & \vdots \\ a_{3,1} & \cdots & a_{i-r,i-r} & & \mathbf{a}_{i-r,(i-r+1)i} & & \cdots & a_{i-r,n-2} & a_{i-r,n-1} & a_{i-r,n} \\ & & & & & & \cdots & a_{i-r+1,n-2} & a_{i-r+1,n-1} & a_{i-r+1,n} \\ \mathbf{a}_{i,i-r} & \cdots & \mathbf{a}_{(i-r+1)i,i-r} & & \mathbf{A}_{(i-r+1)i,(i-r+1)i} & & \cdots & \vdots & \vdots & \vdots \\ & & & & & & \cdots & a_{i,n-2} & a_{i,n-1} & a_{i,n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & a_{m-2,m-2} & a_{m-2,n-1} & a_{m-2,n} \\ 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 & a_{m-1,n-1} & a_{m-1,n} \\ 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & a_{m,n} \end{bmatrix}$$

Repeat step 1 until matrix $\mathbf{A}$ is transformed into the upper triangular matrix $\mathbf{U}$.

$$\mathbf{U} = \mathbf{A_m} = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \cdots & a_{1,i} & a_{1,i+1} & \cdots & a_{1,n} \\ 0 & a_{2,2} & a_{2,3} & \cdots & a_{2,i} & a_{2,i+1} & \cdots & a_{2,n} \\ 0 & 0 & a_{3,3} & \cdots & a_{3,i} & a_{3,i+1} & \cdots & a_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{i,i} & a_{i,i+1} & \cdots & a_{i,n} \\ 0 & 0 & 0 & \cdots & 0 & a_{i+1,i+1} & \cdots & a_{i+1,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \cdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & a_{m,n} \end{bmatrix}$$

Step 2: After the 2nd row is eliminated from the 1st row, the upper and lower triangular matrices can be formed as

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \mathbf{M}_{2,1} & 1 & 0 & \cdots & 0 \\ \mathbf{M}_{3,1} & \mathbf{M}_{3,2} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{M}_{m,1} & \mathbf{M}_{m,2} & \mathbf{M}_{m,3} & \cdots & 1 \end{bmatrix},$$

C. EXAMPLE

Given a matrix $\mathbf{A}$,

$$\mathbf{A} = \left[\begin{array}{cccccccc} a_{1,1} & \vdots & a_{1,2} & a_{1,3} & \vdots & a_{1,4} & \vdots & a_{1,5} \quad a_{1,6} \\ \cdots & \cdots & \cdots & \cdots & \vdots & & \vdots & \\ a_{2,1} & \vdots & a_{2,2} & a_{2,3} & \vdots & a_{2,4} & \vdots & a_{2,5} \quad a_{2,6} \\ a_{3,1} & \vdots & a_{3,2} & a_{3,3} & \vdots & a_{3,4} & \vdots & a_{3,5} \quad a_{3,6} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots & \\ a_{4,1} & & a_{4,2} & a_{4,3} & \vdots & a_{4,4} & \vdots & a_{4,5} \quad a_{4,6} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \quad \cdots \\ a_{5,1} & & a_{5,2} & a_{5,3} & & a_{5,4} & \vdots & a_{5,5} \quad a_{5,6} \\ a_{6,1} & & a_{6,2} & a_{6,3} & & a_{6,4} & \vdots & a_{6,5} \quad a_{6,6} \end{array}\right], \tag{E.1}$$

where $a_{i,j}$ is the matrix element for the i^{th} row and j^{th} column. It is desired to decompose the matrix using the generalized UL algorithm. For this particular example, it also is desired to eliminate columns 5 and 6 together and followed by the elimination of columns 2 and 3 together, as shown by the partitioning in (E.1). The number of columns to be eliminated can be represented in vector form as $\begin{bmatrix} 1 & 2 & 1 & 2 \end{bmatrix}$, where the numbers refer to the block sizes (number of columns to be eliminated together).

Step 1: Eliminate columns 5 and 6

a. Define the following vectors and matrices

$$\mathbf{A}_{56,56} = \begin{bmatrix} a_{5,5} & a_{5,6} \\ a_{6,5} & a_{6,6} \end{bmatrix}$$

$$\mathbf{a}_{56,1} = \begin{bmatrix} a_{5,1} \\ a_{6,1} \end{bmatrix} \qquad \mathbf{a}_{56,2} = \begin{bmatrix} a_{5,2} \\ a_{6,2} \end{bmatrix} \qquad \mathbf{a}_{56,3} = \begin{bmatrix} a_{5,3} \\ a_{6,3} \end{bmatrix} \qquad \mathbf{a}_{56,4} = \begin{bmatrix} a_{5,4} \\ a_{6,4} \end{bmatrix}$$

$$\mathbf{a}_{1,56} = \begin{bmatrix} a_{1,5} & a_{1,6} \end{bmatrix} \qquad \mathbf{a}_{2,56} = \begin{bmatrix} a_{2,5} & a_{2,6} \end{bmatrix} \qquad \mathbf{a}_{3,56} = \begin{bmatrix} a_{3,5} & a_{3,6} \end{bmatrix} \qquad \mathbf{a}_{4,56} = \begin{bmatrix} a_{4,5} & a_{4,6} \end{bmatrix}$$

b. The multipliers for columns 1 through 4 are

$$\mathbf{M}_{56,1} = \mathbf{A}_{56,56}^{-1}\mathbf{a}_{56,1} \qquad \mathbf{M}_{56,2} = \mathbf{A}_{56,56}^{-1}\mathbf{a}_{56,2}$$
$$\mathbf{M}_{56,3} = \mathbf{A}_{56,56}^{-1}\mathbf{a}_{56,3} \qquad \mathbf{M}_{56,4} = \mathbf{A}_{56,56}^{-1}\mathbf{a}_{56,4}$$

c. Eliminating columns 5 and 6 from columns 1 through 4 produces (primes have been added to the modified values, for this example only, to highlight the changes at each step)

$$\mathbf{a}_{56,1}^{'} = \mathbf{a}_{56,1} - \mathbf{A}_{56,56}\mathbf{M}_{56,1} = 0$$
$$a_{1,1}^{'} = a_{1,1} - \mathbf{a}_{1,56}\mathbf{M}_{56,1} \qquad a_{2,1}^{'} = a_{2,1} - \mathbf{a}_{2,56}\mathbf{M}_{56,1}$$
$$a_{3,1}^{'} = a_{3,1} - \mathbf{a}_{3,56}\mathbf{M}_{56,1} \qquad a_{4,1}^{'} = a_{4,1} - \mathbf{a}_{4,56}\mathbf{M}_{56,1}$$

$$\mathbf{a}_{56,2}^{'} = \mathbf{a}_{56,2} - \mathbf{A}_{56,56}\mathbf{M}_{56,2} = 0$$
$$a_{1,2}^{'} = a_{1,2} - \mathbf{a}_{1,56}\mathbf{M}_{56,2} \qquad a_{2,2}^{'} = a_{2,2} - \mathbf{a}_{2,56}\mathbf{M}_{56,2}$$
$$a_{3,2}^{'} = a_{3,2} - \mathbf{a}_{3,56}\mathbf{M}_{56,2} \qquad a_{4,1}^{'} = a_{4,2} - \mathbf{a}_{4,56}\mathbf{M}_{56,2}$$

$$\mathbf{a}_{56,3}^{'} = \mathbf{a}_{56,3} - \mathbf{A}_{56,56}\mathbf{M}_{56,3} = 0$$
$$a_{1,3}^{'} = a_{1,3} - \mathbf{a}_{1,56}\mathbf{M}_{56,3} \qquad a_{2,3}^{'} = a_{2,3} - \mathbf{a}_{2,56}\mathbf{M}_{56,3}$$
$$a_{3,3}^{'} = a_{3,3} - \mathbf{a}_{3,56}\mathbf{M}_{56,3} \qquad a_{4,3}^{'} = a_{4,3} - \mathbf{a}_{4,56}\mathbf{M}_{56,3}$$

$$\mathbf{a}_{56,4}^{'} = \mathbf{a}_{56,4} - \mathbf{A}_{56,56}\mathbf{M}_{56,4} = 0$$
$$a_{1,4}^{'} = a_{1,4} - \mathbf{a}_{1,56}\mathbf{M}_{56,4} \qquad a_{2,4}^{'} = a_{2,4} - \mathbf{a}_{2,56}\mathbf{M}_{56,4}$$
$$a_{3,4}^{'} = a_{3,4} - \mathbf{a}_{3,56}\mathbf{M}_{56,4} \qquad a_{4,4}^{'} = a_{4,4} - \mathbf{a}_{4,56}\mathbf{M}_{56,4}$$

d. Matrix $\mathbf{A}$ with columns 5 and 6 eliminated looks like

$$\mathbf{A}^{'} = \begin{bmatrix} a_{1,1}^{'} & \vdots & a_{1,2}^{'} & a_{1,3}^{'} & \vdots & a_{1,4}^{'} & \vdots & a_{1,5} & a_{1,6} \\ \dots & \dots & \dots & \dots & \vdots & & \vdots & & \\ a_{2,1}^{'} & \vdots & a_{2,2}^{'} & a_{2,3}^{'} & \vdots & a_{2,4}^{'} & \vdots & a_{2,5} & a_{2,6} \\ a_{3,1}^{'} & \vdots & a_{3,2}^{'} & a_{3,3}^{'} & \vdots & a_{3,4}^{'} & \vdots & a_{3,5} & a_{3,6} \\ \dots & \dots & \dots & \dots & \dots & \dots & \vdots & & \\ a_{4,1}^{'} & & a_{4,2}^{'} & a_{4,3}^{'} & \vdots & a_{4,4}^{'} & \vdots & a_{4,5} & a_{4,6} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & & 0 & 0 & & 0 & \vdots & a_{5,5} & a_{5,6} \\ 0 & & 0 & 0 & & 0 & \vdots & a_{6,5} & a_{6,6} \end{bmatrix}.$$

Step 2: Eliminate row 4

a. The multipliers for columns 1 through 3 are

$$M_{4,1} = a_{4,1}^{'}/a_{4,4}^{'} \qquad M_{4,2} = a_{4,2}^{'}/a_{4,4}^{'} \qquad M_{4,3} = a_{4,3}^{'}/a_{4,4}^{'}$$

b. Eliminating column 4 from columns 1 through 3 produces

$$a_{4,1}^{''} = a_{4,1}^{'} - a_{4,4}^{'}M_{4,1} = 0 \qquad a_{3,1}^{''} = a_{3,1}^{'} - a_{3,4}^{'}M_{4,1}$$
$$a_{2,1}^{''} = a_{2,1}^{'} - a_{2,4}^{'}M_{4,1} \qquad a_{1,1}^{''} = a_{1,1}^{'} - a_{1,4}^{'}M_{4,1}$$

$$a_{4,2}^{''} = a_{4,2}^{'} - a_{4,4}^{'}M_{4,2} = 0 \qquad a_{3,2}^{''} = a_{3,2}^{'} - a_{3,4}^{'}M_{4,2}$$
$$a_{2,2}^{''} = a_{2,2}^{'} - a_{2,4}^{'}M_{4,2} \qquad a_{1,2}^{''} = a_{1,2}^{'} - a_{1,4}^{'}M_{4,2}$$

$$a_{4,3}^{''} = a_{4,3}^{'} - a_{4,4}^{'}M_{4,3} = 0 \qquad a_{3,3}^{''} = a_{3,3}^{'} - a_{3,4}^{'}M_{4,3}$$
$$a_{2,3}^{''} = a_{2,3}^{'} - a_{2,4}^{'}M_{4,3} \qquad a_{1,3}^{''} = a_{1,3}^{'} - a_{1,4}^{'}M_{4,3}$$

c. Matrix $\mathbf{A}$ with columns 4 through 6 eliminated looks like

$$\mathbf{A}^{''} = \begin{bmatrix} a^{''}_{1,1} & \vdots & a^{''}_{1,2} & a^{''}_{1,3} & \vdots & a^{'}_{1,4} & \vdots & a_{1,5} & a_{1,6} \\ \cdots & \cdots & \cdots & \cdots & \vdots & & \vdots & & \\ a^{''}_{2,1} & \vdots & a^{''}_{2,2} & a^{''}_{2,3} & \vdots & a^{'}_{2,4} & \vdots & a_{2,5} & a_{2,6} \\ a^{''}_{3,1} & \vdots & a^{''}_{3,2} & a^{''}_{3,3} & \vdots & a^{'}_{3,4} & \vdots & a_{3,5} & a_{3,6} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots & & \\ 0 & & 0 & 0 & \vdots & a^{'}_{4,4} & \vdots & a_{4,5} & a_{4,6} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & & 0 & 0 & & 0 & \vdots & a_{5,5} & a_{5,6} \\ 0 & & 0 & 0 & & 0 & \vdots & a_{6,5} & a_{6,6} \end{bmatrix}.$$

Step 3: Eliminate columns 2 and 3

a. Define the following vectors and matrices

$$\mathbf{A}_{23,23} = \begin{bmatrix} a^{''}_{2,2} & a^{''}_{2,3} \\ a^{''}_{3,2} & a^{''}_{3,3} \end{bmatrix} \qquad \mathbf{a}_{23,1} = \begin{bmatrix} a^{''}_{2,1} \\ a^{''}_{3,1} \end{bmatrix} \qquad \mathbf{a}_{1,23} = \begin{bmatrix} a^{''}_{1,2} & a^{''}_{1,3} \end{bmatrix}$$

b. The multiplier for column 1 is

$$\mathbf{M}_{23,1} = \mathbf{A}^{-1}_{23,23}\mathbf{a}_{23,1}$$

c. Eliminating columns 2 and 3 from column 1 produces

$$\mathbf{a}^{'''}_{23,1} = \mathbf{a}_{23,1} - \mathbf{A}_{23,23}\mathbf{M}_{23,1} = 0 \qquad a^{'''}_{1,1} = a^{''}_{1,1} - \mathbf{a}_{1,23}\mathbf{M}_{23,1}$$

d. Matrix **A** with columns 2 and 3 eliminated looks like

$$\mathbf{A}''' = \begin{bmatrix} a'''_{1,1} & \vdots & a''_{1,2} & a''_{1,3} & \vdots & a'_{1,4} & \vdots & a_{1,5} & a_{1,6} \\ \cdots & \cdots & \cdots & \cdots & \vdots & & \vdots & & \\ 0 & \vdots & a''_{2,2} & a''_{2,3} & \vdots & a'_{2,4} & \vdots & a_{2,5} & a_{2,6} \\ 0 & \vdots & a''_{3,2} & a''_{3,3} & \vdots & a'_{3,4} & \vdots & a_{3,5} & a_{3,6} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots & & \\ 0 & & 0 & 0 & \vdots & a'_{4,4} & \vdots & a_{4,5} & a_{4,6} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & & 0 & 0 & & 0 & \vdots & a_{5,5} & a_{5,6} \\ 0 & & 0 & 0 & & 0 & \vdots & a_{6,5} & a_{6,6} \end{bmatrix}.$$

Step 4: Build the **U** and **L** matrices

a. The matrix **U** is the transformed matrix **A**

$$\mathbf{U} = \begin{bmatrix} a'''_{1,1} & \vdots & a''_{1,2} & a''_{1,3} & \vdots & a'_{1,4} & \vdots & a_{1,5} & a_{1,6} \\ \cdots & \cdots & \cdots & \cdots & \vdots & & \vdots & & \\ 0 & \vdots & a''_{2,2} & a''_{2,3} & \vdots & a'_{2,4} & \vdots & a_{2,5} & a_{2,6} \\ 0 & \vdots & a''_{3,2} & a''_{3,3} & \vdots & a'_{3,4} & \vdots & a_{3,5} & a_{3,6} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots & & \\ 0 & & 0 & 0 & \vdots & a'_{4,4} & \vdots & a_{4,5} & a_{4,6} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & & 0 & 0 & & 0 & \vdots & a_{5,5} & a_{5,6} \\ 0 & & 0 & 0 & & 0 & \vdots & a_{6,5} & a_{6,6} \end{bmatrix}.$$

b. The matrix $\mathbf{L}$ is made up of the multipliers, such that

$$\mathbf{L} = \begin{bmatrix} 1 & \vdots & 0 & 0 & \vdots & 0 & \vdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \vdots & & \vdots & & \\ \mathbf{M}_{23,1} & \vdots & 1 & 0 & \vdots & 0 & \vdots & 0 & 0 \\ & \vdots & 0 & 1 & \vdots & 0 & \vdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots & & \\ M_{4,1} & & M_{4,2} & M_{4,3} & \vdots & 1 & \vdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathbf{M}_{56,1} & & \mathbf{M}_{56,2} & \mathbf{M}_{56,3} & & \mathbf{M}_{56,4} & \vdots & 1 & 0 \\ & & & & & & \vdots & 0 & 1 \end{bmatrix}$$

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

[1] Ram G. Shenoy, Daniel Burnside, and Thomas W. Parks. Linear Periodic Systems and Multirate Filter Design. *IEEE Transactions on Signal Processing*, 42(9):2242–2256, 1994.

[2] Bor-Sen Chen and You-Li Chen. Multirate Modeling of AR/MA Stochastic Signals and Its Application to the Combined Estimation-Interpolation Problem. *IEEE Transactions on Signal Processing*, 43(10):2302–2312, 1995.

[3] Bor-Sen Chen, Chin-Wei Lin, and You-Li Chen. Optimal Signal Reconstruction in Noisy Filter Bank Systems: Multirate Kalman Filtering Approach. *IEEE Transactions on Signal Processing*, 43(11):2496–2504, 1995.

[4] Tsuhan Chen and P. P. Vaidyanathan. The Role of Integer Matrices in Multidimensional Multirate Systems. *IEEE Transactions on Signal Processing*, 41(3):1035–1047, 1993.

[5] Brian L. Evans, Robert H. Bamberger, and James H. McClellan. Rules for Multidimensional Multirate Structures. *IEEE Transactions on Signal Processing*, 42(4):762–771, 1994.

[6] Shuichi Ohno and Hideaki Sakai. Optimization of Filter Banks Using Cyclostationary Spectral Analysis. *IEEE Transactions on Signal Processing*, 44(11):2718–2725, 1996.

[7] Dominick Andrisani, II and Ching-Fau Gau. Estimation Using a Multirate Filter. *IEEE Transactions on Automatic Control*, 32(7):653–656, 1987.

[8] Bor-Sen Chen and Chin-Wei Lin. Optimal Design of Deconvolution Filters for Stochastic Multirate Signal Systems. *Elsevier Signal Processing*, 47:287–305, 1995.

[9] You-Li Chen and Bor-Sen Chen. Model-Based Multirate Representation of Speech Signals and Its Application to Recovery of Missing Speech Packets. *IEEE Transactions on Speech and Audio Processing*, 5(3):220–231, 1997.

[10] P. De Leon, W. Kober, J. Krumvieda, and J. Thomas. Subband Kalman Filtering with Applications to Target Tracking. In *The Tenth International Conference on Signal Processing Applications & Technologies*, Orlando, Florida, 1999. IEEE.

[11] J. Q. Ni, K. L. Ho, K. W. Tse, J. S. Ni, and M. H. Shen. Multirate Kalman Filtering Approach for Optimal Two-Dimensional Signal Reconstruction from Noisy Subband Systems. In *International Conference on Image Processing*, pages 157–160, Santa Barbara, California, 1997. IEEE.

[12] D. Dumur and P. Boucher. New Predictive Technics: Control Axis Solutions. In *Third IEEE Conference on Control Applications*, volume 3, pages 1663–1668, Glasgow, Scotland, UK, 1994. IEEE.

[13] Matteo Rossi and Marco Lovera. A Multirate Predictive Approach to Orbit Control of a Small Spacecraft. In *American Control Conference*, volume 3, pages 1843–1848, Anchorage, Alaska, 2002. IEEE.

[14] Josep Tornero, Ricardo Piza, Pedro Albertos, and Julian Salt. Multirate LQG Controller Applied to Self-Location and Path-Tracking in Mobile Robots. In *International Conference on Intelligent Robots and Systems*, pages 625–630, Maui, Hawaii, USA, 2001. IEEE.

[15] Roberto Cristi and Murali Tummala. Multirate, Multiresolution, Recursive Kalman Filter. *Elsevier Signal Processing*, 80:1945–1958, 2000.

[16] Ronald E. Crochiere and Lawrence R. Rabiner. *Multirate Digital Signal Processing*. Prentice Hall, New Jersey, 1983.

[17] P. P. Vaidyanathan. *Multirate Systems and Filter Banks*. Prentice Hall, New Jersey, 1993.

[18] Charles W. Therrien. *Discrete Random Signals and Statistical Signal Processing*. Signal Processing Series. Prentice Hall, New Jersey, 1992.

[19] Louis L. Scharf. *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Electrical and Computer Engineering: Digital Signal Processing. Addision-Wesley Publishing Company, Inc., Menlo Park, California, 1991.

[20] Abraham Wald. *Sequential Analysis*. John Wiley & Sons, Inc., New York, 1947.

[21] Charles W. Therrien. A Sequential Approach to Target Discrimination. *IEEE Transactions on Aerospace and Electronic Systems*, pages 433–440, May 1978.

[22] Charles W. Therrien. *Decision, Estimation, and Classification: An Introduction to Pattern Recognition and Related Topics*. John Wiley & Sons, New York, 1989.

[23] Aryan Saadat Mehr and Tongwen Chen. Properties of Linear Switching Time-Varying Discrete-Time Systems with Applications. *Elsevier Systems & Control Letters*, 39:229–235, 2000.

[24] Aryan Saadat Mehr and Tongwen Chen. Representations of Linear Periodically Time-Varying and Multirate Systems. *IEEE Transactions on Signal Processing*, 50(9):2221–2229, 2002.

[25] See-May Phoong and P. P. Vaidyanathan. Time-Varying Filters and Filter Banks: Some Basic Principles. *IEEE Transactions on Signal Processing*, 44(12):2971–2987, 1996.

[26] See-May Phoong and P. P. Vaidyanathan. Factorability of Lossless Time-Varying Filters and Filter Banks. *IEEE Transactions on Signal Processing*, 45(8):1971–1986, 1997.

[27] P. P. Vaidyanathan and Ahmet Kirac. Cyclic LTI Systems in Digital Signal Processing. *IEEE Transactions on Signal Processing*, 47(2):433–447, 1999.

[28] Vikram M. Gadre and R. K. Patney. Vector Multirate Filtering and Matrix Filter Banks. In *1992 IEEE International Symposium on Circuits and Systems*, volume 3, pages 1360–1363, San Diego, California, 1992. IEEE.

[29] Mark S. Spurbeck and Louis L. Scharf. Least Squares Filter Design for Periodically Correlated Time Series. In *IEEE Seventh SP Workshop on Statistical Signal and Array Processing*, pages 267–270, Quebec, Ontario, Canada, 1994. IEEE.

[30] Pradeep Misra. Time-Invariant Representations of Discrete Periodic Systems. *Automatica*, 32:267–272, 1996.

[31] William A. Gardner. Spectral Characterization of N-th Order Cyclostationarity. In *Fifth ASSP Workshop on Spectrum Estimation and Modeling*, pages 251–255, Rochester, New York, 1990. IEEE.

[32] William A. Gardner. Exploitation of Spectral Redundancy in Cyclostationary Signals. *IEEE Signal Processing Magazine*, 8(2):14–36, 1991.

[33] Sony Akkarakaran and P. P. Vaidyanathan. Bifrequency and Bispectrum Maps: A New Look at Multirate Systems with Stochastic Inputs. *IEEE Transactions on Signal Processing*, 48(3):723–736, 2000.

[34] Charles W. Therrien. Issues in Multirate Statistical Signal Processing. In *Proceedings of the 35th Asilomar Conference on Signals, Systems, and Computers*, volume 1, pages 573–576, Pacific Grove, California, 2001. IEEE.

[35] Charles W. Therrien. Defining Correlation Functions and Power Spectra for Multirate Random Processes. In *Proceedings of the IEEE International Symposium*

on Circuits and Systems (ISCAS), 2005, pages 6010–6013, Kobe, Japan, 2005. IEEE.

[36] Charles W. Therrien. Some Considerations for Statistical Characterization of Nonstationary Random Processes. In *Proceedings of the 36th Asilomar Conference on Signals, Systems, and Computers*, volume 2, pages 1554–1558, Pacific Grove, California, 2002. IEEE.

[37] Charles W. Therrien and Roberto Cristi. Two-dimensional Spectral Representation of Periodic, Cyclostationary, and More General Random Processes. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 4, pages 3561–3563, Orlando, Florida, 2002. IEEE.

[38] Antonio Napolitano. Cyclic Higher-Order Statistics: Input/Output Relations for Discrete- and Continuous-Time MIMO Linear Almost-Periodically Time-Variant Systems. *Elsevier Signal Processing*, 42:147–166, 1995.

[39] Luciano Izzo and Antonio Napolitano. Multirate Processing of Time Series Exhibiting Order Cyclostationarity. *IEEE Transactions on Signal Processing*, 46(2):429–439, 1998.

[40] Arda Yurdakul and Gunhan Dundar. Statistical Methods for the Estimation of Quantization Effects in FIR-Based Multirate Systems. *IEEE Transactions on Signal Processing*, 47(6):1749–1753, 1999.

[41] Vinay P. Sathe and P. P. Vaidyanathan. Effects of Multirate Systems on the Statistical Properties of Random Signals. *IEEE Transactions on Signal Processing*, 41(1):131–146, 1993.

[42] B. Sridhar, P. Smith, R. Suorsa, and B. Hussein. Multirate and Event Driven Kalman Filters for Helicopter Passive Ranging. In *First IEEE Conference on Control Applications*, volume 2, pages 800–805, Dayton, Ohio, 1992. IEEE.

[43] B. Sridhar, P. Smith, R. Suorsa, and B. Hussien. Multirate and Event-Driven Kalman Filters for Helicopter Flight. *IEEE Control Systems Magazine*, pages 26–33, August 1993.

[44] Michele Basseville, Albert Benveniste, Kenneth C. Chou, Stuart A. Golden, Ramine Nikoukhah, and Alan S. Willsky. Modeling and Estimation of Multiresolution Stochastic Processes. *IEEE Transactions on Information Theory*, 38(2):766–784, 1992.

[45] Kenneth C. Chou, Alan S. Willsky, and Albert Benveniste. Multiscale Recursive Estimation, Data Fusion, and Regularization. *IEEE Transactions on Automatic Control*, 39(3):464–478, 1994.

[46] Kenneth C. Chou, Alan S. Willsky, and Ramine Nikoukhah. Multiscale Systems, Kalman Filters, and Riccati Equations. *IEEE Transactions on Automatic Control*, 39(3):479–492, 1994.

[47] Wassim M. Haddad, Dennis S. Bernstein, and Hsing-Hsin Huang. Reduced-Order Multirate Estimation for Stable and Unstable Plants. In *29th Conference on Decision and Control*, pages 2892–2897, Honolulu, Hawaii, 1990. IEEE.

[48] Lang Hong. Multirate Estimation. In *National Aerospace and Electronics Conference*, pages 435–440, Dayton, Ohio, 1994. IEEE.

[49] Lang Hong. Approximating Multirate Estimation Using a Filter Bank. In *33rd Conference on Decision and Control*, Lake Buena Vista, Florida, 1994. IEEE.

[50] James W. Scrofani. *Theory of Multirate Signal Processing with Application to One- and Two-Dimensional Signal Reconstruction.* Doctor of philosophy, Naval Postgraduate School, 2005.

[51] Roberto Cristi, D. A. Koupatsiaris, and Charles W. Therrien. Multirate Filtering and Estimation: The Multirate Wiener Filter. In *Proceedings of the 34th Asilo-*

mar Conference on Signals, Systems, and Computers, volume 1, pages 450–454, Pacific Grove, California, 2000. IEEE.

[52] Ryan J. Kuchler and Charles W. Therrien. Optimal Filtering with Multirate Observations. In *Proceedings of the 37th Asilomar Conference on Signals, Systems, and Computers*, volume 1, pages 1208–1212, Pacific Grove, California, 2003. IEEE.

[53] Charles W. Therrien. Multirate Linear Prediction. In *Proceedings of the European Signal Processing Conference (EUSIPCO)*, Antalya, Turkey, 2005. European Association for Signal, Speech and Image Processing (EUROSIP).

[54] Anthony H. Hawes and Charles W. Therrien. LMS Adaptive Filtering with Multirate Observations. In *Proceedings of the 37th Asilomar Conference on Signals, Systems, and Computers*, volume 1, pages 567–570, Pacific Grove, California, 2003. IEEE.

[55] Charles W. Therrien and Anthony H. Hawes. Least Squares Optimal Filtering with Multirate Observations. In *Proceedings of the 36th Asilomar Conference on Signals, Systems, and Computers*, volume 2, pages 1782–1786, Pacific Grove, California, 2002. IEEE.

[56] James W. Scrofani and Charles W. Therrien. A Stochastic Multirate Signal Processing Approach to High-Resolution Signal Reconstruction. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 4, pages 561–564, Philadelphia, Pennsylvania, 2005. IEEE.

[57] James W. Scrofani and Charles W. Therrien. A Multirate Approach to High-Resolution Image Reconstruction. In *Proceedings of the Seventh IASTED International Conference on Signal and Image Processing*, Honolulu, Hawaii, 2005. International Association of Science and Technology for Development (IASTED).

[58] John G. Proakis and Dimitris G. Manolakis. *Digital Signal Processing: Principles. Algorithms, and Applications*. Macmillan Publishing Company, New York, second edition, 1992.

[59] John William Scott Cassels. *An Introduction to the Geometry of Numbers*. Springer-Verlag, New York, 1971.

[60] Yao Wang, Jörn Ostermann, and Ya-Qin Zhang. *Video Processing and Communications*. Prentice Hall, New Jersey, 2002.

[61] Dimitrios Koupatsiaris. *Analysis of Multirate Random Signals*. Engineer's, Naval Postgraduate School, December 2000.

[62] Alexander Graham. *Kronecker Products and Matrix Calculus: with Applications*. Ellis Horwood Limited, Chichester, England, 1981.

[63] John W. Brewer. Kronecker Products and Matrix Calculus in System Theory. *IEEE Transactions on Circuits and Systems*, 25(9):772–781, 1978.

[64] R. A. Wiggins and E. A. Robinson. Recursive Solutions to the Multichannel Filtering Problem. *Journal of Geophysical Research*, 70:1885–1891, 1965.

[65] Gilbert Strang. *Linear Algebra and Its Applications*. Harcourt Brace Jovanovich, Inc., San Diego, third edition, 1988.

[66] Gene H. Golub and Alexander F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, Maryland, third edition, 1996.

[67] Keinosuke Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, New York, second edition, 1990.

[68] Harry L. Van Trees. *Detection, Estimation, and Modulation Theory - Part I*. John Wiley & Sons, New York, 1968.

[69] Thomas Kailath. *Linear Systems*. Prentice Hall, Inc., Englewood Cliffs, New Jersey, 1980.

[70] S. Lawrence Marple, Jr. *Digital Spectral Analysis with Applications*. Prentice Hall, Inc., Englewood Cliffs, New Jersey, 1987.

[71] Steven M. Kay. *Modern Spectral Estimation: Theory and Application*. Prentice Hall, Inc., Englewood Cliffs, New Jersey, 1988.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Fort Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Dr. Jeffrey B. Knorr, Code EC/Ko
   Chairman, Department of Electrical and Computer Engineering
   Naval Postgraduate School
   Monterey, California

4. Dr. Charles W. Therrien, Code EC/Ti
   Department of Electrical and Computer Engineering
   Naval Postgraduate School
   Monterey, California

5. Dr. Carlos Borges, Code MA/Bc
   Department of Mathematics
   Naval Postgraduate School
   Monterey, California

6. Dr. Roberto Cristi, Code EC/Cx
   Department of Electrical and Computer Engineering
   Naval Postgraduate School
   Monterey, California

7. Dr. Anthony Healey, Code ME/Hy
   Department of Mechanical and Astronautical Engineering
   Naval Postgraduate School
   Monterey, California

8. Dr. Murali Tummala, Code EC/Tu
   Department of Electrical and Computer Engineering
   Naval Postgraduate School
   Monterey, California

9. LCDR Ryan J. Kuchler
   Naval Postgraduate School
   Monterey, California